

1 Reilly T. Stoler (SBN 310761)
2 LIEFF CABRASER HEIMANN & BERNSTEIN, LLP
3 275 Battery Street, 29th Floor
4 San Francisco, CA 94111-3339
Telephone: (415) 956-1000
rstoler@lchb.com

5 Rohit D. Nath (SBN 316062)
6 SUSMAN GODFREY L.L.P
7 1900 Avenue of the Stars, Suite 1400
8 Los Angeles, CA 90067-2906
Telephone: (310) 789-3100
RNath@susmangodfrey.com

9 *Attorneys for Plaintiffs and the Proposed Class*

10
11 UNITED STATES DISTRICT COURT
12 NORTHERN DISTRICT OF CALIFORNIA
13 SAN FRANCISCO DIVISION

14
15 *In re* MOTION TO COMPEL
16 COMPLIANCE WITH NON-PARTY
SUBPOENAS

17 In the matter of

18 AUTHORS GUILD, *et al.*, individually
and on behalf of others similarly situated,

19 Plaintiffs,

20 v.

21 OPENAI INC., *et al.*,

22 Defendants.

Misc. Case No. 3:25-mc-80017

Underlying Litigations: *Authors Guild, et al. v.*
OpenAI, et al., 1:23-cv-08292-SHS, S.D.N.Y.

**DECLARATION OF REILLY STOLER IN
SUPPORT OF MOTION TO COMPEL
DARIO AMODEI AND BENJAMIN MANN
TO TESTIFY AT DEPOSITIONS**

1 I, Reilly T. Stoler, declare as follows:

2 1. I am an attorney licensed to practice in the State of California, and I am a member
3 in good standing of the State Bar of California. I am also admitted to practice in the federal courts
4 of the Northern District of California.

5 2. I am a partner at the law firm Lief, Cabraser, Heimann & Bernstein, LLP.
6 Alongside Susman Godfrey, LLP and Cowan, DeBaets, Abrahams, & Sheppard, LLP, my law
7 firm serves as Interim Co-Lead Counsel for the Fiction and Nonfiction Author Classes. *Authors*
8 *Guild, et al. v. OpenAI, et al.*; 1:23-cv-08292, Dkt. 70.

9 3. I make this Declaration in Support of the consolidated *Authors Guild v. OpenAI*
10 Plaintiffs' Motion to Compel Dario Amodei and Benjamin Mann to Testify at Depositions and of
11 my own personal knowledge. If called upon to testify, I could and would testify competently to
12 the truth of the matters stated herein.

13 4. Attached as **Exhibit A** is a true and correct copy of the subpoenas Plaintiffs served
14 on Dario Amodei and Benjamin Mann seeking Deposition Testimony and Production of
15 Documents pursuant to Rule 45 of the Federal Rules of Civil Procedure.

16 5. Attached as **Exhibit B** is a true and correct copy of the operative Complaint in
17 *Authors Guild, et al. v. OpenAI, et al.*; 1:23-cv-08292 (S.D.N.Y.).

18 6. Attached as **Exhibit C** is a true and correct copy of a March 22, 2024 letter from
19 OpenAI Counsel, Joseph C. Gratz to Plaintiffs' Counsel, stating in relevant part, [REDACTED]

20 [REDACTED]

21 [REDACTED]

22 7. Attached as **Exhibit D** is a true and correct copy of an April 14, 2024 email from
23 OpenAI Counsel Allyson Bennett to Plaintiffs' Counsel, [REDACTED]

24 [REDACTED]

25 [REDACTED]

26 8. Attached as **Exhibit E** is a true and correct copy of a June 14, 2024 email from
27 Amodei and Mann's Counsel, Catherine Malone, stating in relevant part her "view that [Amodei
28 and Mann] should be deposed once across the OpenAI cases."

1 9. Attached as **Exhibit F** is a true and correct copy of an August 28, 2024 email from
2 me to Counsel for Amodei and Mann recording the topics discussed in the conferral earlier that
3 same day, and stating in relevant part that Amodei and Mann would not provide convenient dates
4 for deposition until the entry of a deposition coordination order in the *Authors Guild, et al. v.*
5 *OpenAI, et al.* litigation.

6 10. Attached as **Exhibit G** is a true and correct copy of a January 7, 2025 email from
7 me to Amodei and Mann’s Counsel, reflecting the discussions and agreements reached the
8 conferral earlier that same day, and stating in relevant part that Counsel for Amodei and Mann
9 “agreed to relay convenient deposition dates for Messrs. Amodei and Mann, which will take place
10 in the Bay Area.”

11 11. Attached as **Exhibit H** is a true and correct copy of a January 16, 2025 email from
12 Amodei and Mann’s Counsel, Catherine Malone, stating in relevant part that (1) “Mr. Amodei is
13 not available for deposition,” and (2) “Mr. Mann can be available of a four-hour deposition,
14 coordinated with the plaintiffs in the *In re OpenAI ChatGPT Litigation*, and after [Plaintiffs in
15 *Authors Guild v. OpenAI* litigation] complete substantially all of the OpenAI depositions [we]
16 intend on taking. This offer is contingent on an agreement not to depose Mr. Amodei.”

17 12. Attached as **Exhibit I** is a true and correct copy of a January 17, 2025 email from
18 me to Counsel for Amodei and Mann, stating in relevant part that Plaintiffs are willing “to
19 negotiat[e] reasonable accommodations on date, location, duration and coordination with *In re*
20 *ChatGPT*.” In response, neither Amodei nor Mann changed their position as to these subpoenas.

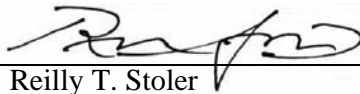
21 13. On January 16, 2025, I conferred with Counsel for Amodei and Mann, Catherine
22 Malone and Todd Scott, regarding completed and forthcoming document productions from
23 Amodei and Mann as well as the timing of their depositions. The conferral lasted over an hour.
24 Ms. Malone and Mr. Scott cited Mr. Amodei’s busy work schedule as Chief Executive Officer of
25 Anthropic P.B.C. as the sole basis for his unavailability to testify at a deposition. Ms. Malone and
26 Mr. Scott cited Mann’s obligations as the parent of a young child and a family member’s medical
27 diagnosis as the bases for his unavailability to testify at a deposition.

1 14. Plaintiffs' Subpoenas to Amodei and Mann contemplate that the depositions will
2 take place in San Francisco. I have made clear to Counsel for Amodei and Mann that Plaintiffs
3 are willing to negotiate on the location, including at Amodei and Mann's respective offices at
4 normal business hours.

5 15. Attached as **Exhibit J** is a true and correct copy of Plaintiffs' First Set of
6 Interrogatories served on OpenAI on February 23, 2024, seeking in relevant part the five
7 individuals who are most knowledgeable about the creation and contents of the Books2 dataset.

8 16. Attached as **Exhibit K** is a true and correct copy of OpenAI's Supplemental
9 Response to Interrogatory No. 1, served on April 29, 2024, [REDACTED]

10 [REDACTED].
11
12 I declare under penalty of perjury that the forgoing is true and correct. Executed in San
13 Francisco, California, this 28th day of January 2025.

14
15 
16 _____

Reilly T. Stoler

17 LIEFF CABRASER HEIMANN & BERNSTEIN, LLP
18 275 Battery Street, 29th Floor
19 San Francisco, CA 94111-3339
20 rstoler@lchb.com
21 Telephone: 415.956.1000
22 Facsimile: 415.956.1008
23
24
25
26
27
28

EXHIBIT A

AO 88A (Rev. 12/13) Subpoena to Testify at a Deposition in a Civil Action

UNITED STATES DISTRICT COURT

for the

Southern District of New York

Authors Guild et al.

Plaintiff

v.

OpenAI, Inc. et al.

Defendant

Civil Action No. 1:23-cv-08292; 1:23-cv-10211

SUBPOENA TO TESTIFY AT A DEPOSITION IN A CIVIL ACTION

To:

Dario Amodei

(Name of person to whom this subpoena is directed)

☒ **Testimony:** **YOU ARE COMMANDED** to appear at the time, date, and place set forth below to testify at a deposition to be taken in this civil action. If you are an organization, you must designate one or more officers, directors, or managing agents, or designate other persons who consent to testify on your behalf about the following matters, or those set forth in an attachment:

Place: 275 Battery Street, 29th Floor San Francisco, CA 94111	Date and Time: 05/20/2024 3:00 pm
---	--------------------------------------

The deposition will be recorded by this method: audiovisual and stenographic.

☒ **Production:** You, or your representatives, must also bring with you to the deposition the following documents, electronically stored information, or objects, and must permit inspection, copying, testing, or sampling of the material: See Attachment A.

The following provisions of Fed. R. Civ. P. 45 are attached – Rule 45(c), relating to the place of compliance; Rule 45(d), relating to your protection as a person subject to a subpoena; and Rule 45(e) and (g), relating to your duty to respond to this subpoena and the potential consequences of not doing so.

Date: 04/29/2024

CLERK OF COURT

OR

/s/ Wesley J. DozierSignature of Clerk or Deputy ClerkAttorney's signature

The name, address, e-mail address, and telephone number of the attorney representing (name of party) Counsel for Fiction and Nonfiction Plaintiffs and the Proposed Classes, who issues or requests this subpoena, are:

Wesley J. Dozier, Lieff, Cabraser, Heimann, and Bernstein, LLP, 222 2nd Ave. S., Nashville, TN 37206. Rachel Geman, Lieff, Cabraser, Heimann, and Bernstein, LLP, 250 Hudson Street, New York, NY 10013.

Notice to the person who issues or requests this subpoena

If this subpoena commands the production of documents, electronically stored information, or tangible things, a notice and a copy of the subpoena must be served on each party in this case before it is served on the person to whom it is directed. Fed. R. Civ. P. 45(a)(4).

AO 88A (Rev. 12/13) Subpoena to Testify at a Deposition in a Civil Action (Page 2)

Civil Action No. 1:23-cv-08292; 1:23-cv-10211

PROOF OF SERVICE*(This section should not be filed with the court unless required by Fed. R. Civ. P. 45.)*

I received this subpoena for *(name of individual and title, if any)* _____
 on *(date)* _____ .

☐ I served the subpoena by delivering a copy to the named individual as follows: _____
 _____ on *(date)* _____ ; or

☐ I returned the subpoena unexecuted because: _____
 _____ .

Unless the subpoena was issued on behalf of the United States, or one of its officers or agents, I have also
 tendered to the witness the fees for one day's attendance, and the mileage allowed by law, in the amount of
 \$ _____ .

My fees are \$ _____ for travel and \$ _____ for services, for a total of \$ 0.00 .

I declare under penalty of perjury that this information is true.

Date: _____

Server's signature

Printed name and title

Server's address

Additional information regarding attempted service, etc.:

Federal Rule of Civil Procedure 45 (c), (d), (e), and (g) (Effective 12/1/13)**(c) Place of Compliance.**

(1) For a Trial, Hearing, or Deposition. A subpoena may command a person to attend a trial, hearing, or deposition only as follows:

- (A) within 100 miles of where the person resides, is employed, or regularly transacts business in person; or
- (B) within the state where the person resides, is employed, or regularly transacts business in person, if the person
 - (i) is a party or a party's officer; or
 - (ii) is commanded to attend a trial and would not incur substantial expense.

(2) For Other Discovery. A subpoena may command:

- (A) production of documents, electronically stored information, or tangible things at a place within 100 miles of where the person resides, is employed, or regularly transacts business in person; and
- (B) inspection of premises at the premises to be inspected.

(d) Protecting a Person Subject to a Subpoena; Enforcement.

(1) Avoiding Undue Burden or Expense; Sanctions. A party or attorney responsible for issuing and serving a subpoena must take reasonable steps to avoid imposing undue burden or expense on a person subject to the subpoena. The court for the district where compliance is required must enforce this duty and impose an appropriate sanction—which may include lost earnings and reasonable attorney's fees—on a party or attorney who fails to comply.

(2) Command to Produce Materials or Permit Inspection.

(A) *Appearance Not Required.* A person commanded to produce documents, electronically stored information, or tangible things, or to permit the inspection of premises, need not appear in person at the place of production or inspection unless also commanded to appear for a deposition, hearing, or trial.

(B) *Objections.* A person commanded to produce documents or tangible things or to permit inspection may serve on the party or attorney designated in the subpoena a written objection to inspecting, copying, testing, or sampling any or all of the materials or to inspecting the premises—or to producing electronically stored information in the form or forms requested. The objection must be served before the earlier of the time specified for compliance or 14 days after the subpoena is served. If an objection is made, the following rules apply:

- (i) At any time, on notice to the commanded person, the serving party may move the court for the district where compliance is required for an order compelling production or inspection.
- (ii) These acts may be required only as directed in the order, and the order must protect a person who is neither a party nor a party's officer from significant expense resulting from compliance.

(3) Quashing or Modifying a Subpoena.

(A) *When Required.* On timely motion, the court for the district where compliance is required must quash or modify a subpoena that:

- (i) fails to allow a reasonable time to comply;
- (ii) requires a person to comply beyond the geographical limits specified in Rule 45(c);
- (iii) requires disclosure of privileged or other protected matter, if no exception or waiver applies; or
- (iv) subjects a person to undue burden.

(B) *When Permitted.* To protect a person subject to or affected by a subpoena, the court for the district where compliance is required may, on motion, quash or modify the subpoena if it requires:

(i) disclosing a trade secret or other confidential research, development, or commercial information; or

(ii) disclosing an unretained expert's opinion or information that does not describe specific occurrences in dispute and results from the expert's study that was not requested by a party.

(C) *Specifying Conditions as an Alternative.* In the circumstances described in Rule 45(d)(3)(B), the court may, instead of quashing or modifying a subpoena, order appearance or production under specified conditions if the serving party:

- (i) shows a substantial need for the testimony or material that cannot be otherwise met without undue hardship; and
- (ii) ensures that the subpoenaed person will be reasonably compensated.

(e) Duties in Responding to a Subpoena.

(1) Producing Documents or Electronically Stored Information. These procedures apply to producing documents or electronically stored information:

(A) *Documents.* A person responding to a subpoena to produce documents must produce them as they are kept in the ordinary course of business or must organize and label them to correspond to the categories in the demand.

(B) *Form for Producing Electronically Stored Information Not Specified.* If a subpoena does not specify a form for producing electronically stored information, the person responding must produce it in a form or forms in which it is ordinarily maintained or in a reasonably usable form or forms.

(C) *Electronically Stored Information Produced in Only One Form.* The person responding need not produce the same electronically stored information in more than one form.

(D) *Inaccessible Electronically Stored Information.* The person responding need not provide discovery of electronically stored information from sources that the person identifies as not reasonably accessible because of undue burden or cost. On motion to compel discovery or for a protective order, the person responding must show that the information is not reasonably accessible because of undue burden or cost. If that showing is made, the court may nonetheless order discovery from such sources if the requesting party shows good cause, considering the limitations of Rule 26(b)(2)(C). The court may specify conditions for the discovery.

(2) Claiming Privilege or Protection.

(A) *Information Withheld.* A person withholding subpoenaed information under a claim that it is privileged or subject to protection as trial-preparation material must:

- (i) expressly make the claim; and
- (ii) describe the nature of the withheld documents, communications, or tangible things in a manner that, without revealing information itself privileged or protected, will enable the parties to assess the claim.

(B) *Information Produced.* If information produced in response to a subpoena is subject to a claim of privilege or of protection as trial-preparation material, the person making the claim may notify any party that received the information of the claim and the basis for it. After being notified, a party must promptly return, sequester, or destroy the specified information and any copies it has; must not use or disclose the information until the claim is resolved; must take reasonable steps to retrieve the information if the party disclosed it before being notified; and may promptly present the information under seal to the court for the district where compliance is required for a determination of the claim. The person who produced the information must preserve the information until the claim is resolved.

(g) Contempt.

The court for the district where compliance is required—and also, after a motion is transferred, the issuing court—may hold in contempt a person who, having been served, fails without adequate excuse to obey the subpoena or an order related to it.

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

AUTHORS GUILD, *et al.*, individually and on
behalf of others similarly situated,

Plaintiffs,

v.

OPENAI INC., *et al.*,

Defendants.

ECF CASE

No. 1:23-cv-08292-SHS

JONATHAN ALTER, *et al.*, on behalf of
themselves and all others similarly situated,

Plaintiffs,

v.

OPENAI, INC., *et al.*,

Defendants.

No. 1:23-cv-10211-SHS

ATTACHMENT A TO DOCUMENT SUBPOENA TO DARIO AMODEI

Pursuant to Rule 45 of the Federal Rules of Civil Procedure, Plaintiffs in the above-captioned civil action command **You** to attend a deposition, the testimony of which will be recorded by audiovisual and stenographic means. Further, **You** are commanded to bring with you to the deposition or produce before the date of the deposition the materials described below, at the location shown on the attached subpoena. Plaintiffs are also amenable to an electronic production.

DEFINITIONS

1. **Action** means the above captioned litigations, *Authors Guild, et. al., v. OpenAI, et. al.*, No. 1:23-cv-08292-SHS (S.D.N.Y.) and *Jonathan Alter, et al., v. Open AI Inc. et al.*, No. 1:23-cv-10211-SHS (S.D.N.Y.).

2. **Books1 and/or Books2** means either of two internet-based books corpora known by the name **Books1** and **Books2**, respectively, that were to train the **LLMs** that underlie **ChatGPT**.

3. **ChatGPT** means all consumer-facing versions of the chatbot application OpenAI first released in November 2022 and any LLM underlying those consumer-facing applications.

4. **Common Crawl** means an internet-based books corpora known as **Common Crawl** that was used train the **LLMs** that underlie **ChatGPT**.

5. **Communication** means the conveyance (in the form of facts, ideas, thoughts, opinions, data, inquiries or otherwise) of information and **Includes** correspondence, memoranda, reports, presentations, face-to-face conversations, telephone conversations, text messages, instant messages, voice messages, negotiations, agreements, inquiries, understandings, meetings, letters, notes, mail, email, exchanges of recorded information, and postings of any type. **Communications Includes** instances where one party disseminates information that the other party receives but does not respond to.

6. **Concern** or **Concerning** means to be the subject of, make reference to, comment on, discuss, describe, identify, or contain text or images about the stated topic.

7. **Documents** means all materials within the full scope of Federal Rule of Civil Procedure 34 **Including** all writings and recordings, **Including** the originals, drafts and all non-

identical copies, whether different from the original by reason of any notation made on such copies or otherwise (**Including** email and attachments, correspondence, memoranda, notes, diaries, statistics, letters, telegrams, minutes, contracts, reports, studies, checks, statements, tags, labels, invoices, brochures, periodicals, receipts, returns, summaries, pamphlets, books, interoffice and intra-office communications, instant messages, chats, offers, notations of any sort of conversations, working papers, applications, permits, file wrappers, indices, telephone calls, meetings or printouts, teletypes, telefax, invoices, worksheets, and all drafts, alterations, modifications, changes and amendments of any of the foregoing), graphic or aural representations of any kind (**Including** photographs, charts, microfiche, microfilm, videotape, recordings, motion pictures, plans, drawings, surveys), and electronic, mechanical, magnetic, optical, or electric records or representations of any kind (**Including** computer files and programs, tapes, cassettes, discs, and recordings), **Including** metadata.

8. **Including** means including but not limited to.

9. **Large Language Model** or **LLM** means deep learning algorithm that can perform a variety of natural language processes, trained on large datasets, including, but not limited to, ChatGPT, GPT-3.5, and GPT-4.

10. **Microsoft** means Microsoft Corporation, any of their directors, officers, employees, partners, members, representatives, agents (**Including** attorneys, accountants, consultants, investment advisors or bankers), and any other person acting or purporting to act on their behalf, as well as corporate parents, subsidiaries, affiliates, predecessor entities, successor entities, divisions, departments, groups, acquired entities, related entities, or any other entity acting or purporting to act on their behalf.

11. **OpenAI** means OpenAI, Inc., OpenAI GP, LLC, OpenAI, LLC, OpenAI OPCO LLC, OpenAI Global LLC, OAI Corporation, LLC, OpenAI Holdings, LLC, and any of their directors, officers, employees, partners, members, representatives, agents (**Including** attorneys, accountants, consultants, investment advisors or bankers), and any other person acting or purporting to act on their behalf, as well as corporate parents, subsidiaries, affiliates, predecessor entities, successor entities, divisions, departments, groups, acquired entities, related entities, or any other entity acting or purporting to act on their behalf.

12. **Relate to** or **Relating** means concerning, referring to, describing, evidencing or constituting, in whole or in part.

13. **You** and **Your** means Dario Amodei and any of Mr. Amodei's directors, officers, employees, partners, members, representatives, agents (**Including** attorneys, accountants, consultants, investment advisors or bankers), and any other person acting or purporting to act on Mr. Amodei's behalf.

INSTRUCTIONS

1. Please produce each **Document** requested herein in its entirety and without deletion or excisions, regardless of whether **You** consider the entire document to be relevant or responsive.

2. If any **Document** is known to have existed but no longer exists, has been destroyed, or is otherwise unavailable, please identify the **Document**, the reason for its loss, destruction, or unavailability, the name of each person known or reasonably believed by **You** to have had possession, custody, or control of the original and any copy thereof (if applicable), and a description of the disposition of each copy of the **Document**.

3. If no **Documents** responsive to a particular request exist, please state that no responsive **Documents** exist.

4. Construe the conjunctions “and” and “or” non-restrictively or nonexclusively if doing so would bring within the scope of these Requests **Documents** that might otherwise be construed to be outside of their scope.

5. Construe the use of the singular to include the plural; the use of the masculine, feminine, or neuter gender to include the others; and the use of one form of the verb to include the others if doing so would bring within the scope of these Requests **Documents** that might otherwise be construed to be outside of their scope.

DOCUMENT REQUESTS

Request No. 1: Documents sufficient to determine **Your** employment at **OpenAI**, including your job title, responsibilities, the date range of your employment, and the reason(s) for your departure.

Request No. 2: All Documents and Communications regarding **Books1 and/or Books2, including**, but not limited to, **Communications** regarding the sources used to compile **Books1 and/or Books2**.

Request No. 3: Documents sufficient to determine **Your** knowledge of the creation, maintenance, and/or use of **Books1 and/or Books2**.

Request No. 4: All Documents and Communications regarding **Common Crawl, including**, but not limited to, **Communications** regarding the sources used to compile **Common Crawl**.

Request No. 5: All Documents and Communications sufficient to determine **Your** knowledge of the creation, maintenance, and/or use of any other dataset by **OpenAI** that may contain copyrighted material.

Request No. 6: All Documents and Communications Related to any complaints or concerns you raised with **OpenAI** regarding the development of the **LLMs** that support **ChatGPT**, **including**, but not limited to, complaints raised concerning **OpenAI's** alleged copyright infringement.

Request No. 7: All Documents and Communications between **You** and **Microsoft**, **including**, but not limited to **Documents** and **Communications** sufficient to determine **Microsoft's** involvement with the development of **ChatGPT**.

Dated: April 29, 2024

/s/ Wesley J. Dozier

Rachel Geman
LIEFF CABRASER HEIMANN & BERNSTEIN, LLP
250 Hudson Street, 8th Floor
New York, New York 10013-1413
Tel.: 212-355-9500
rgeman@lchb.com

Reilly T. Stoler (*pro hac vice*)
LIEFF CABRASER HEIMANN & BERNSTEIN, LLP
275 Battery Street, 29th Floor
San Francisco, CA 94111-3339
Tel.: 415-956-1000
rstoler@lchb.com

Wesley Dozier (*pro hac vice*)
LIEFF CABRASER HEIMANN & BERNSTEIN, LLP
222 2nd Avenue, Suite 1640
Nashville, TN 37201
Tel.: 615-313-9000
wdozier@lchb.com

Justin A. Nelson (*pro hac vice*)

Alejandra C. Salinas (*pro hac vice*)
SUSMAN GODFREY L.L.P.
1000 Louisiana Street, Suite 5100
Houston, TX 77002
Tel.: 713-651-9366
jnelson@susmangodfrey.com
asalinas@susmangodfrey.com

Rohit D. Nath (*pro hac vice*)
SUSMAN GODFREY L.L.P.
1900 Avenue of the Stars, Suite 1400
Los Angeles, California 90067
Tel.: 310-789-3100
rnath@susmangodfrey.com

J. Craig Smyser
SUSMAN GODFREY L.L.P.
1901 Avenue of the Americas, 32nd Floor
New York, New York 10019
Tel.: 212-336-8330
csmyser@susmangodfrey.com

Scott J. Shoulder
CeCe M. Cole
COWAN DEBAETS ABRAHAMS & SHEPPARD LLP
41 Madison Avenue, 38th Floor
New York, New York 10010
Tel.: 212-974-7474
sshoulder@cdas.com
ccole@cdas@com

Attorney for Plaintiffs and the Proposed Classes

AO 88A (Rev. 12/13) Subpoena to Testify at a Deposition in a Civil Action

UNITED STATES DISTRICT COURT

for the

Southern District of New York

Authors Guild et al.

Plaintiff

v.

OpenAI, Inc. et al.

Defendant

Civil Action No. 1:23-cv-08292; 1:23-cv-10211

SUBPOENA TO TESTIFY AT A DEPOSITION IN A CIVIL ACTION

To:

Dario Amodei

(Name of person to whom this subpoena is directed)

☒ **Testimony:** **YOU ARE COMMANDED** to appear at the time, date, and place set forth below to testify at a deposition to be taken in this civil action. If you are an organization, you must designate one or more officers, directors, or managing agents, or designate other persons who consent to testify on your behalf about the following matters, or those set forth in an attachment:

Place: 275 Battery Street, 29th Floor San Francisco, CA 94111	Date and Time: 05/20/2024 3:00 pm
---	--------------------------------------

The deposition will be recorded by this method: audiovisual and stenographic.

☒ **Production:** You, or your representatives, must also bring with you to the deposition the following documents, electronically stored information, or objects, and must permit inspection, copying, testing, or sampling of the material: See Attachment A.

The following provisions of Fed. R. Civ. P. 45 are attached – Rule 45(c), relating to the place of compliance; Rule 45(d), relating to your protection as a person subject to a subpoena; and Rule 45(e) and (g), relating to your duty to respond to this subpoena and the potential consequences of not doing so.

Date: 04/29/2024

CLERK OF COURT

OR

/s/ Wesley J. DozierSignature of Clerk or Deputy ClerkAttorney's signature

The name, address, e-mail address, and telephone number of the attorney representing (name of party) Counsel for
Fiction and Nonfiction Plaintiffs and the Proposed Classes, who issues or requests this subpoena, are:

Wesley J. Dozier, Lieff, Cabraser, Heimann, and Bernstein, LLP, 222 2nd Ave. S., Nashville, TN 37206. Rachel Geman, Lieff, Cabraser, Heimann, and Bernstein, LLP, 250 Hudson Street, New York, NY 10013.

Notice to the person who issues or requests this subpoena

If this subpoena commands the production of documents, electronically stored information, or tangible things, a notice and a copy of the subpoena must be served on each party in this case before it is served on the person to whom it is directed. Fed. R. Civ. P. 45(a)(4).

AO 88A (Rev. 12/13) Subpoena to Testify at a Deposition in a Civil Action (Page 2)

Civil Action No. 1:23-cv-08292; 1:23-cv-10211

PROOF OF SERVICE

(This section should not be filed with the court unless required by Fed. R. Civ. P. 45.)

I received this subpoena for *(name of individual and title, if any)* _____
on *(date)* _____.

☐ I served the subpoena by delivering a copy to the named individual as follows: _____
_____ on *(date)* _____; or

☐ I returned the subpoena unexecuted because: _____
_____.

Unless the subpoena was issued on behalf of the United States, or one of its officers or agents, I have also
tendered to the witness the fees for one day's attendance, and the mileage allowed by law, in the amount of
\$ _____.

My fees are \$ _____ for travel and \$ _____ for services, for a total of \$ 0.00 .

I declare under penalty of perjury that this information is true.

Date: _____

Server's signature

Printed name and title

Server's address

Additional information regarding attempted service, etc.:

Federal Rule of Civil Procedure 45 (c), (d), (e), and (g) (Effective 12/1/13)**(c) Place of Compliance.**

(1) For a Trial, Hearing, or Deposition. A subpoena may command a person to attend a trial, hearing, or deposition only as follows:

- (A) within 100 miles of where the person resides, is employed, or regularly transacts business in person; or
- (B) within the state where the person resides, is employed, or regularly transacts business in person, if the person
 - (i) is a party or a party's officer; or
 - (ii) is commanded to attend a trial and would not incur substantial expense.

(2) For Other Discovery. A subpoena may command:

- (A) production of documents, electronically stored information, or tangible things at a place within 100 miles of where the person resides, is employed, or regularly transacts business in person; and
- (B) inspection of premises at the premises to be inspected.

(d) Protecting a Person Subject to a Subpoena; Enforcement.

(1) Avoiding Undue Burden or Expense; Sanctions. A party or attorney responsible for issuing and serving a subpoena must take reasonable steps to avoid imposing undue burden or expense on a person subject to the subpoena. The court for the district where compliance is required must enforce this duty and impose an appropriate sanction—which may include lost earnings and reasonable attorney's fees—on a party or attorney who fails to comply.

(2) Command to Produce Materials or Permit Inspection.

(A) *Appearance Not Required.* A person commanded to produce documents, electronically stored information, or tangible things, or to permit the inspection of premises, need not appear in person at the place of production or inspection unless also commanded to appear for a deposition, hearing, or trial.

(B) *Objections.* A person commanded to produce documents or tangible things or to permit inspection may serve on the party or attorney designated in the subpoena a written objection to inspecting, copying, testing, or sampling any or all of the materials or to inspecting the premises—or to producing electronically stored information in the form or forms requested. The objection must be served before the earlier of the time specified for compliance or 14 days after the subpoena is served. If an objection is made, the following rules apply:

- (i) At any time, on notice to the commanded person, the serving party may move the court for the district where compliance is required for an order compelling production or inspection.
- (ii) These acts may be required only as directed in the order, and the order must protect a person who is neither a party nor a party's officer from significant expense resulting from compliance.

(3) Quashing or Modifying a Subpoena.

(A) *When Required.* On timely motion, the court for the district where compliance is required must quash or modify a subpoena that:

- (i) fails to allow a reasonable time to comply;
- (ii) requires a person to comply beyond the geographical limits specified in Rule 45(c);
- (iii) requires disclosure of privileged or other protected matter, if no exception or waiver applies; or
- (iv) subjects a person to undue burden.

(B) *When Permitted.* To protect a person subject to or affected by a subpoena, the court for the district where compliance is required may, on motion, quash or modify the subpoena if it requires:

(i) disclosing a trade secret or other confidential research, development, or commercial information; or

(ii) disclosing an unretained expert's opinion or information that does not describe specific occurrences in dispute and results from the expert's study that was not requested by a party.

(C) *Specifying Conditions as an Alternative.* In the circumstances described in Rule 45(d)(3)(B), the court may, instead of quashing or modifying a subpoena, order appearance or production under specified conditions if the serving party:

- (i) shows a substantial need for the testimony or material that cannot be otherwise met without undue hardship; and
- (ii) ensures that the subpoenaed person will be reasonably compensated.

(e) Duties in Responding to a Subpoena.

(1) Producing Documents or Electronically Stored Information. These procedures apply to producing documents or electronically stored information:

(A) *Documents.* A person responding to a subpoena to produce documents must produce them as they are kept in the ordinary course of business or must organize and label them to correspond to the categories in the demand.

(B) *Form for Producing Electronically Stored Information Not Specified.* If a subpoena does not specify a form for producing electronically stored information, the person responding must produce it in a form or forms in which it is ordinarily maintained or in a reasonably usable form or forms.

(C) *Electronically Stored Information Produced in Only One Form.* The person responding need not produce the same electronically stored information in more than one form.

(D) *Inaccessible Electronically Stored Information.* The person responding need not provide discovery of electronically stored information from sources that the person identifies as not reasonably accessible because of undue burden or cost. On motion to compel discovery or for a protective order, the person responding must show that the information is not reasonably accessible because of undue burden or cost. If that showing is made, the court may nonetheless order discovery from such sources if the requesting party shows good cause, considering the limitations of Rule 26(b)(2)(C). The court may specify conditions for the discovery.

(2) Claiming Privilege or Protection.

(A) *Information Withheld.* A person withholding subpoenaed information under a claim that it is privileged or subject to protection as trial-preparation material must:

- (i) expressly make the claim; and
- (ii) describe the nature of the withheld documents, communications, or tangible things in a manner that, without revealing information itself privileged or protected, will enable the parties to assess the claim.

(B) *Information Produced.* If information produced in response to a subpoena is subject to a claim of privilege or of protection as trial-preparation material, the person making the claim may notify any party that received the information of the claim and the basis for it. After being notified, a party must promptly return, sequester, or destroy the specified information and any copies it has; must not use or disclose the information until the claim is resolved; must take reasonable steps to retrieve the information if the party disclosed it before being notified; and may promptly present the information under seal to the court for the district where compliance is required for a determination of the claim. The person who produced the information must preserve the information until the claim is resolved.

(g) Contempt.

The court for the district where compliance is required—and also, after a motion is transferred, the issuing court—may hold in contempt a person who, having been served, fails without adequate excuse to obey the subpoena or an order related to it.

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

AUTHORS GUILD, *et al.*, individually and on
behalf of others similarly situated,

Plaintiffs,

v.

OPENAI INC., *et al.*,

Defendants.

ECF CASE

No. 1:23-cv-08292-SHS

JONATHAN ALTER, *et al.*, on behalf of
themselves and all others similarly situated,

Plaintiffs,

v.

OPENAI, INC., *et al.*,

Defendants.

No. 1:23-cv-10211-SHS

ATTACHMENT A TO DOCUMENT SUBPOENA TO BENJAMIN J. MANN

Pursuant to Rule 45 of the Federal Rules of Civil Procedure, Plaintiffs in the above-captioned civil action command **You** to attend a deposition, the testimony of which will be recorded by audiovisual and stenographic means. Further, **You** are commanded to bring with you to the deposition or produce before the date of the deposition the materials described below, at the location shown on the attached subpoena. Plaintiffs are also amenable to an electronic production.

DEFINITIONS

1. **Action** means the above captioned litigations, *Authors Guild, et. al., v. OpenAI, et. al.*, No. 1:23-cv-08292-SHS (S.D.N.Y.) and *Jonathan Alter, et al., v. Open AI Inc. et al.*, No. 1:23-cv-10211-SHS (S.D.N.Y.).

2. **Books1 and/or Books2** means either of two internet-based books corpora known by the name **Books1** and **Books2**, respectively, that were to train the **LLMs** that underlie **ChatGPT**.

3. **ChatGPT** means all consumer-facing versions of the chatbot application OpenAI first released in November 2022 and any LLM underlying those consumer-facing applications.

4. **Common Crawl** means an internet-based books corpora known as **Common Crawl** that was used train the **LLMs** that underlie **ChatGPT**.

5. **Communication** means the conveyance (in the form of facts, ideas, thoughts, opinions, data, inquiries or otherwise) of information and **Includes** correspondence, memoranda, reports, presentations, face-to-face conversations, telephone conversations, text messages, instant messages, voice messages, negotiations, agreements, inquiries, understandings, meetings, letters, notes, mail, email, exchanges of recorded information, and postings of any type. **Communications Includes** instances where one party disseminates information that the other party receives but does not respond to.

6. **Concern** or **Concerning** means to be the subject of, make reference to, comment on, discuss, describe, identify, or contain text or images about the stated topic.

7. **Documents** means all materials within the full scope of Federal Rule of Civil Procedure 34 **Including** all writings and recordings, **Including** the originals, drafts and all non-

identical copies, whether different from the original by reason of any notation made on such copies or otherwise (**Including** email and attachments, correspondence, memoranda, notes, diaries, statistics, letters, telegrams, minutes, contracts, reports, studies, checks, statements, tags, labels, invoices, brochures, periodicals, receipts, returns, summaries, pamphlets, books, interoffice and intra-office communications, instant messages, chats, offers, notations of any sort of conversations, working papers, applications, permits, file wrappers, indices, telephone calls, meetings or printouts, teletypes, telefax, invoices, worksheets, and all drafts, alterations, modifications, changes and amendments of any of the foregoing), graphic or aural representations of any kind (**Including** photographs, charts, microfiche, microfilm, videotape, recordings, motion pictures, plans, drawings, surveys), and electronic, mechanical, magnetic, optical, or electric records or representations of any kind (**Including** computer files and programs, tapes, cassettes, discs, and recordings), **Including** metadata.

8. **Including** means including but not limited to.

9. **Large Language Model** or **LLM** means deep learning algorithm that can perform a variety of natural language processes, trained on large datasets, including, but not limited to, ChatGPT, GPT-3.5, and GPT-4.

10. **Microsoft** means Microsoft Corporation, any of their directors, officers, employees, partners, members, representatives, agents (**Including** attorneys, accountants, consultants, investment advisors or bankers), and any other person acting or purporting to act on their behalf, as well as corporate parents, subsidiaries, affiliates, predecessor entities, successor entities, divisions, departments, groups, acquired entities, related entities, or any other entity acting or purporting to act on their behalf.

11. **OpenAI** means OpenAI, Inc., OpenAI GP, LLC, OpenAI, LLC, OpenAI OPCO LLC, OpenAI Global LLC, OAI Corporation, LLC, OpenAI Holdings, LLC, and any of their directors, officers, employees, partners, members, representatives, agents (**Including** attorneys, accountants, consultants, investment advisors or bankers), and any other person acting or purporting to act on their behalf, as well as corporate parents, subsidiaries, affiliates, predecessor entities, successor entities, divisions, departments, groups, acquired entities, related entities, or any other entity acting or purporting to act on their behalf.

12. **Relate to** or **Relating** means concerning, referring to, describing, evidencing or constituting, in whole or in part.

13. **You** and **Your** means Benjamin J. Mann and any of Mr. Mann's directors, officers, employees, partners, members, representatives, agents (**Including** attorneys, accountants, consultants, investment advisors or bankers), and any other person acting or purporting to act on Mr. Mann's behalf.

INSTRUCTIONS

1. Please produce each **Document** requested herein in its entirety and without deletion or excisions, regardless of whether **You** consider the entire document to be relevant or responsive.

2. If any **Document** is known to have existed but no longer exists, has been destroyed, or is otherwise unavailable, please identify the **Document**, the reason for its loss, destruction, or unavailability, the name of each person known or reasonably believed by **You** to have had possession, custody, or control of the original and any copy thereof (if applicable), and a description of the disposition of each copy of the **Document**.

3. If no **Documents** responsive to a particular request exist, please state that no responsive **Documents** exist.

4. Construe the conjunctions “and” and “or” non-restrictively or nonexclusively if doing so would bring within the scope of these Requests **Documents** that might otherwise be construed to be outside of their scope.

5. Construe the use of the singular to include the plural; the use of the masculine, feminine, or neuter gender to include the others; and the use of one form of the verb to include the others if doing so would bring within the scope of these Requests **Documents** that might otherwise be construed to be outside of their scope.

DOCUMENT REQUESTS

Request No. 1: Documents sufficient to determine **Your** employment at **OpenAI**, including your job title, responsibilities, the date range of your employment, and the reason(s) for your departure.

Request No. 2: All Documents and Communications regarding **Books1 and/or Books2, including**, but not limited to, **Communications** regarding the sources used to compile **Books1 and/or Books2**.

Request No. 3: Documents sufficient to determine **Your** knowledge of the creation, maintenance, and/or use of **Books1 and/or Books2**.

Request No. 4: All Documents and Communications regarding **Common Crawl, including**, but not limited to, **Communications** regarding the sources used to compile **Common Crawl**.

Request No. 5: All Documents and Communications sufficient to determine **Your** knowledge of the creation, maintenance, and/or use of any other dataset by **OpenAI** that may contain copyrighted material.

Request No. 6: All Documents and Communications Related to any complaints or concerns you raised with **OpenAI** regarding the development of the **LLMs** that support **ChatGPT**, **including**, but not limited to, complaints raised concerning **OpenAI's** alleged copyright infringement.

Request No. 7: All Documents and Communications between **You** and **Microsoft**, **including**, but not limited to **Documents** and **Communications** sufficient to determine **Microsoft's** involvement with the development of **ChatGPT**.

Dated: April 29, 2024

/s/ Wesley J. Dozier

Rachel Geman
LIEFF CABRASER HEIMANN & BERNSTEIN, LLP
250 Hudson Street, 8th Floor
New York, New York 10013-1413
Tel.: 212-355-9500
rgeman@lchb.com

Reilly T. Stoler (*pro hac vice*)
LIEFF CABRASER HEIMANN & BERNSTEIN, LLP
275 Battery Street, 29th Floor
San Francisco, CA 94111-3339
Tel.: 415-956-1000
rstoler@lchb.com

Wesley Dozier (*pro hac vice*)
LIEFF CABRASER HEIMANN & BERNSTEIN, LLP
222 2nd Avenue, Suite 1640
Nashville, TN 37201
Tel.: 615-313-9000
wdozier@lchb.com

Justin A. Nelson (*pro hac vice*)

Alejandra C. Salinas (*pro hac vice*)
SUSMAN GODFREY L.L.P.
1000 Louisiana Street, Suite 5100
Houston, TX 77002
Tel.: 713-651-9366
jnelson@susmangodfrey.com
asalinas@susmangodfrey.com

Rohit D. Nath (*pro hac vice*)
SUSMAN GODFREY L.L.P.
1900 Avenue of the Stars, Suite 1400
Los Angeles, California 90067
Tel.: 310-789-3100
rnath@susmangodfrey.com

J. Craig Smyser
SUSMAN GODFREY L.L.P.
1901 Avenue of the Americas, 32nd Floor
New York, New York 10019
Tel.: 212-336-8330
csmyser@susmangodfrey.com

Scott J. Shoulder
CeCe M. Cole
COWAN DEBAETS ABRAHAMS & SHEPPARD LLP
41 Madison Avenue, 38th Floor
New York, New York 10010
Tel.: 212-974-7474
sshoulder@cdas.com
ccole@cdas@com

Attorney for Plaintiffs and the Proposed Classes

EXHIBIT B

**UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK**

AUTHORS GUILD, DAVID BALDACCI,
MARY BLY, MICHAEL CONNELLY, SYLVIA
DAY, JONATHAN FRANZEN, JOHN
GRISHAM, ELIN HILDERBRAND,
CHRISTINA BAKER KLINE, MAYA
SHANBHAG LANG, VICTOR LAVALLE,
GEORGE R.R. MARTIN, JODI PICOULT,
DOUGLAS PRESTON, ROXANA ROBINSON,
GEORGE SAUNDERS, SCOTT TUROW, and
RACHEL VAIL, individually and on behalf of
others similarly situated,

Plaintiffs,

v.

OPEN AI INC., OPENAI OPCO LLC, OPENAI
GP LLC, OPENAI, LLC, OPENAI GLOBAL
LLC, OAI CORPORATION LLC, OPENAI
HOLDINGS LLC, OPENAI STARTUP FUND I
LP, OPENAI STARTUP FUND GP I LLC,
OPENAI STARTUP FUND MANAGEMENT
LLC, and MICROSOFT CORPORATION,

Defendants.

JONATHAN ALTER, KAI BIRD, TAYLOR
BRANCH, RICH COHEN, EUGENE LINDEN,
DANIEL OKRENT, JULIAN SANCTON,
HAMPTON SIDES, STACY SCHIFF, JAMES
SHAPIRO, JIA TOLENTINO, and SIMON
WINCHESTER, on behalf of themselves and all
others similarly situated,

Plaintiffs,

v.

OPEN AI INC., OPENAI OPCO LLC, OPENAI
GP LLC, OPENAI, LLC, OPENAI GLOBAL
LLC, OAI CORPORATION LLC, OPENAI
HOLDINGS LLC, OPENAI STARTUP FUND I

ECF CASE

No. 1:23-cv-08292-SHS;
No. 1:23-cv-10211-SHS

**FIRST CONSOLIDATED CLASS
ACTION COMPLAINT**

JURY TRIAL DEMANDED

LP, OPENAI STARTUP FUND GP I LLC,
OPENAI STARTUP FUND MANAGEMENT
LLC, and MICROSOFT CORPORATION,

Defendants.

Plaintiffs Jonathan Alter, The Authors Guild, David Baldacci, Kai Bird, Mary Bly, Taylor Branch, Rich Cohen, Michael Connelly, Sylvia Day, Jonathan Franzen, John Grisham, Elin Hilderbrand, Christina Baker Kline, Maya Shanbhag Lang, Victor LaValle, Eugene Linden, George R.R. Martin, Daniel Okrent, Jodi Picoult, Douglas Preston, Roxana Robinson, Julian Sancton, George Saunders, Stacy Schiff, Hampton Sides, James Shapiro, Jia Tolentino, Scott Turow, Simon Winchester, and Rachel Vail, on behalf of themselves and all other similarly situated (the “Fiction Author Class” and “Nonfiction Author Class,” as defined below), for their complaint against Defendants OpenAI, Inc., OpenAI GP LLC, OpenAI, LLC, OpenAI OpCo LLC, OpenAI Global LLC, OAI Corporation, LLC, OpenAI Holdings, LLC, OpenAI Startup Fund I LP, OpenAI Startup Fund GP I LLC, OpenAI Startup Fund Management LLC (collectively “OpenAI”), and Microsoft Corporation (all collectively “Defendants”), allege as follows:

NATURE OF THE CASE

1. OpenAI and Microsoft have built a business valued into the tens of billions of dollars by taking the combined works of humanity without permission. Rather than pay for intellectual property (such as when a person buys a book), they choose to operate as if the laws protecting copyright do not exist. Yet the United States Constitution protects the fundamental principle that creators, like authors, deserve compensation for their works, and the Copyright Act grants “a bundle of exclusive rights” to creators, including “the rights to reproduce the

copyrighted work[s].” *Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508, 526 (2023).

2. Plaintiffs, authors of a broad array of works of fiction and nonfiction, bring this action under the Copyright Act seeking redress for Defendants’ flagrant and harmful infringements of Plaintiffs’ registered copyrights. Defendants copied Plaintiffs’ works and then fed them into their “large language models” or “LLMs,” algorithms designed to output human-seeming text responses to users’ prompts and queries. These algorithms are at the heart of Defendants’ massive commercial enterprise. And at the heart of these algorithms is systematic theft on a mass scale.

3. Defendants’ LLMs endanger authors’ ability to make a living, in that the LLMs allow anyone to generate—automatically and freely (or very cheaply)—texts that they would otherwise pay writers to create. Moreover, the LLMs can spit out derivative works: material that is based on, mimics, summarizes, or paraphrases Plaintiffs’ works, and harms the market for them. Without Plaintiffs’ copyrighted works on which to “train” their LLMs, Defendants would have no commercial product with which to damage—if not usurp—the market for these professional writers’ works. OpenAI’s willful copying thus makes Plaintiffs’ works into engines of Plaintiffs’ own destruction, notwithstanding that professional writers often spend years conceiving of and writing their creations.

4. Specifically, Defendants OpenAI and Microsoft collaborated closely to create and monetize the generative artificial intelligence models known as GPT-3, GPT-3.5, GPT-4, and GPT-4 Turbo. These are the computer models that power the popular ChatGPT chatbot, which Defendants have followed with a suite of other commercial offerings, like ChatGPT Enterprise, ChatGPT Plus, Bing Chat, Browse with Bing, Microsoft Copilot, and others. Defendants’ GPT

models have been designed to recognize and process text inputs from a user and, in response, generate text that has been calibrated to mimic a human written response.

5. That end product—a computer model and chatbot built to mimic human written expression—came at a price. Defendants’ models were “trained,” in Defendants’ parlance, by reproducing a massive corpus of copyrighted material, including, upon information and belief, tens or hundreds of thousands of fiction and nonfiction books. The only way that Defendants’ models could be trained to generate text output that resembles human expression is to copy and analyze a large, diverse corpus of text written by humans. In training their models, Defendants reproduced copyrighted texts to exploit precisely what the Copyright Act was designed to protect: the elements of protectible expression within them, like the choice and order of words and sentences, syntax, flow, themes, and paragraph and story structure. In OpenAI’s words, the goal of the training process was to teach their model to “learn” “how words fit together grammatically,” “how words work together to form higher-level ideas,” and “how sequences of words form structured thoughts.”¹ In other words, by training its models on certain works, OpenAI copied the works’ expression so that the models could memorize, mimic, and paraphrase that expression.

6. Defendants copied and data-mined the works of writers, without permission or compensation, to build a machine that is capable (or, as technology advances, will soon be capable) of performing the same type of work for which these writers would be paid. Without the wide corpus of copyrighted material on which to feed, there would be no ChatGPT. Defendants’ commercial success was possible only because they copied and digested the

¹ Fred von Lohmann, response to Notice of Inquiry and Request for Comment 5, (Oct. 30, 2023), *available at* https://downloads.regulations.gov/COLC-2023-0006-8906/attachment_1.pdf.

protected, copyrightable expression contained in billions of pages of actual text, across millions of copyrighted works—all without paying a penny to the authors of those works.

7. OpenAI could have “trained” its LLMs on works in the public domain. It could have paid a reasonable licensing fee to use copyrighted works. What Defendants could *not* do was evade the Copyright Act altogether to power their lucrative commercial endeavor, taking whatever datasets of relatively recent books they could get their hands on without authorization. There is nothing fair about this. OpenAI’s unauthorized use of Plaintiffs’ copyrighted works thus presents a straightforward infringement case applying well-established law to well-recognized copyright harms.

8. OpenAI’s chief executive Sam Altman has told Congress that he shares Plaintiffs’ concerns. According to Altman, “Ensuring that the creator economy continues to be vibrant is an important priority for OpenAI. ... OpenAI does not want to replace creators. We want our systems to be used to empower creativity, and to support and augment the essential humanity of artists and creators.”² Altman testified that OpenAI “think[s] that creators deserve control over how their creations are used” and that “content creators, content owners, need to benefit from this technology.”³ Altman also has represented that OpenAI has “licens[ed] content directly from content owners” for “training” purposes.⁴ Not so from Plaintiffs. As to them, Altman and Defendants have proved unwilling to turn these words into actions.

² Sam Altman, *Questions for the Record*, at 9–10 (June 22, 2023), available at https://www.judiciary.senate.gov/imo/media/doc/2023-05-16_-_qfr_responses_-_altman.pdf (last accessed Dec. 4, 2023).

³ *Oversight of A.I.: Rules for Artificial Intelligence: Hearing Before the S. Judiciary Comm. Subcomm. on Privacy, Tech. and the Law*, 118th Cong. (2023) (testimony of OpenAI CEO Sam Altman), available at <https://techpolicy.press/transcript-senate-judiciary-subcommittee-hearing-on-oversight-of-ai> (last accessed Dec. 4, 2023).

⁴ Altman, *Questions for the Record*, *supra*, at 10.

9. Defendants OpenAI and Microsoft have enjoyed enormous financial gain from their free exploitation of copyrighted material. OpenAI recently reported that it is “generating revenue at a pace of \$1.3 billion a year.”⁵ Microsoft, for its part, has seen its investment in OpenAI increase many-fold and its own GPT-based products, like BingChat, succeed in the marketplace. And its stock price has increased as Microsoft has touted its ability to exploit and leverage AI across its products. Analysts project that the integration of GPT into Microsoft products could generate more than \$10 billion in annualized revenue by 2026,⁶ with just one version of this integration—“GitHub Copilot”—already generating more than \$100 million in annual recurring revenue.⁷ In developing and monetizing these AI products, Microsoft and OpenAI have been close partners every step of the way, from the training of GPT-3 to today.

10. OpenAI and Microsoft’s commercial gain has come at the expense of creators and rightsholders like Plaintiffs and members of the Classes. A person who reads a book typically buys it from a store. But Defendants did not even do that. Neither OpenAI nor Microsoft have paid for the books used to train their models. Nor have Defendants sought to obtain—or pay for—a license to copy and exploit the protected expression contained in the copyrighted works used to train their models. Instead, Defendants took these works; they made unlicensed copies of them; and they used those unlicensed copies to digest and analyze the copyrighted expression in them, all for commercial gain. The end result is a computer model that is not only built on the work of thousands of creators and authors, but also built to generate a wide range of

⁵ AJ Hess, *The Biggest Challenges Facing OpenAI’s Mira Murati, the Newly Minted Most Powerful Woman in Tech*, FAST COMPANY (last visited Nov. 20, 2023), <https://www.fastcompany.com/90985829/the-biggest-challenges-facing-openais-mira-murati-the-newly-minted-most-powerful-woman-in-tech>.

⁶ Jordan Novet, *Microsoft Starts Selling AI Toll for Office, Which Could Generate \$10 Billion a Year by 2026*, CNBC (last visited Nov. 20, 2023) <https://www.cnbc.com/2023/11/01/microsoft-365-copilot-becomes-generally-available.html>.

⁷ Aaron Holmes, *Microsoft’s GitHub AI Coding Assistant Exceeds \$100 Million in Recurring Revenue*, THE INFORMATION, (last visited Nov. 20, 2023), <https://www.theinformation.com/briefings/microsoft-github-copilot-revenue-100-million-ARR-ai> (available at <https://perma.cc/5S7F-4GBY>).

expression—from shortform articles to book chapters—that mimics the syntax, voice, and themes of the copyrighted works on which it was trained.

11. Plaintiffs seek to represent Classes of fiction and nonfiction writers whose works were used to train Defendants’ artificial intelligence models. Plaintiffs, on behalf of themselves and the Classes, seek damages from Defendants for their largescale infringement of their copyrighted works, as well as injunctive relief.

JURISDICTION & VENUE

12. The Court has subject matter jurisdiction under 28 U.S.C. §§ 1331 and 1338(a) because this action arises under the Copyright Act of 1976, 17 U.S.C. § 101, *et seq.*

13. The Court also has personal jurisdiction over Defendants because they have purposely availed themselves of the privilege of conducting business in New York.

14. OpenAI and Microsoft’s copyright infringement and contributory copyright infringement occurred, in substantial part, in this District. OpenAI sold and distributed, and continues to sell and distribute, its GPT products, including ChatGPT, ChatGPT Enterprise, ChatGPT Plus, Browse with Bing, and application programming interface tools (API) within New York and to New York residents. OpenAI marketed and sold GPT-based products to New York residents and New York-based companies.

15. Microsoft distributed and sold GPT-based products, like Bing Chat and Azure products that incorporate GPT-3 and GPT-4. Upon information and belief, Microsoft also assisted OpenAI’s copyright infringement from New York, including from Azure datacenters located in New York, which were used to facilitate OpenAI’s use and exploitation of the training dataset used for development of OpenAI’s GPT models. Microsoft maintains offices and employs personnel in New York. Upon information and belief, Microsoft’s New York personnel were involved in the creation and maintenance of the supercomputing systems that

powered OpenAI's widespread infringement, as well as in the commercialization of OpenAI's GPT models.

16. Plaintiffs The Authors Guild, Bird, Bly, Kline, Lang, LaValle, Linden, Robinson, Sancton, Schiff, Shapiro, Tolentino, Vail, and Winchester are citizens of New York. The injuries alleged here from Defendants' infringement occurred in this District.

17. Venue is proper under 28 U.S.C. § 1400(a) because Defendants or their agents reside or may be found in this District due to their infringing activities, along with their commercialization of their infringing activities, that occurred in this District. Venue is also proper under 28 U.S.C. § 1391(b)(2) because a substantial part of the events giving rise to Plaintiffs' claims occurred in this District, including the sales of Defendants' GPT-based products within this District.

THE PARTIES

A. Plaintiffs

18. Plaintiff The Authors Guild is a nonprofit 501(c)(6) organization based in New York, New York.

19. Plaintiff Jonathan Alter is an author and a resident of New Jersey.

20. Plaintiff David Baldacci is an author and a resident of Virginia.

21. Plaintiff Kai Bird is an author and a resident of New York.

22. Plaintiff Mary Bly is an author and a resident of New York.

23. Plaintiff Taylor Branch is an author and a resident of Maryland.

24. Plaintiff Rich Cohen is an author and a resident of Connecticut.

25. Plaintiff Michael Connelly is an author and a resident of Florida.

26. Plaintiff Sylvia Day is an author and a resident of Nevada.

27. Plaintiff Jonathan Franzen is an author and a resident of California.

28. Plaintiff John Grisham is an author and a resident of Virginia.
29. Plaintiff Elin Hilderbrand is an author and a resident of Massachusetts.
30. Plaintiff Christina Baker Kline is an author and a resident of New York.
31. Plaintiff Maya Shanbhag Lang is an author and a resident of New York.
32. Plaintiff Victor LaValle is an author and a resident of New York.
33. Plaintiff Eugene Linden is an author and a resident of New York.
34. Plaintiff George R.R. Martin is an author and a resident of New Mexico.
35. Plaintiff Daniel Okrent is an author and a resident of Massachusetts.
36. Plaintiff Jodi Picoult is an author and a resident of New Hampshire.
37. Plaintiff Douglas Preston is an author and a resident of New Mexico.
38. Plaintiff Roxana Robinson is an author and a resident of New York.
39. Plaintiff Julian Sancton is an author and a resident of New York.
40. Plaintiff George Saunders is an author and a resident of California.
41. Plaintiff Stacy Schiff is an author and a resident of New York.
42. Plaintiff Hampton Sides is an author and a resident of New Mexico.
43. Plaintiff James Shapiro is an author and a resident of New York.
44. Plaintiff Jia Tolentino is an author and a resident of New York.
45. Plaintiff Scott Turow is an author and a resident of Florida.
46. Plaintiff Simon Winchester OBE is an author and a resident of New York.
47. Plaintiff Rachel Vail is an author and a resident of New York.
48. The registration information for the infringed works of Plaintiffs is identified in

Exhibits A and B to this Complaint.

B. OpenAI Defendants

49. Defendant OpenAI, Inc. is a Delaware nonprofit corporation with a principal place of business in San Francisco, California. OpenAI, Inc. was formed in December 2015. OpenAI, Inc. owns and controls all other OpenAI entities.

50. Defendant OpenAI GP, LLC is a Delaware limited liability company with a principal place of business in San Francisco, California. OpenAI GP, LLC wholly owns and controls OpenAI OpCo LLC, which until recently was known as OpenAI LP. OpenAI, Inc. uses OpenAI GP LLC to control OpenAI OpCo LLC and OpenAI Global, LLC. OpenAI GP LLC was involved in the copyright infringement alleged here through its direction and control of OpenAI OpCo LLC and OpenAI Global LLC.

51. Defendant OpenAI OpCo LLC is a Delaware limited liability company with a principal place of business in San Francisco, California. OpenAI OpCo LLC was formerly known as OpenAI LP. OpenAI OpCo LLC is the sole member of OpenAI, LLC, and has been directly involved in OpenAI's mass infringement and has directed this infringement through its control of OpenAI, LLC. OpenAI OpCo LLC serves as the for-profit arm of OpenAI.

52. Defendant OpenAI, LLC is a Delaware limited liability company with a principal place of business in San Francisco, California. OpenAI, LLC was formed in September 2020. OpenAI LLC monetizes and distributes OpenAI's GPT-based products, all of which born out of OpenAI's copyright infringement. Upon information and belief, OpenAI, LLC is owned and controlled by both OpenAI, Inc. and Microsoft Corporation, through OpenAI Global LLC and OpenAI OpCo LLC.

53. Defendant OpenAI Global LLC is a Delaware limited liability company with a principal place of business in San Francisco, California. Microsoft Corporation has a minority stake in OpenAI Global LLC and OpenAI, Inc. has a majority stake in OpenAI Global LLC, indirectly through OpenAI Holdings LLC and OAI Corporation, LLC. OpenAI Global LLC was

involved in the copyright infringement alleged here through its ownership, control, and direction of OpenAI LLC.

54. Defendant OAI Corporation, LLC is a Delaware limited liability company with a principal place of business in San Francisco, California. OAI Corporation, LLC's sole member is OpenAI Holdings, LLC. OAI Corporation, LLC was and is involved in the unlawful conduct alleged herein through its ownership, control, and direction of OpenAI Global LLC and OpenAI LLC.

55. Defendant OpenAI Holdings, LLC is a Delaware limited liability company, whose sole members are OpenAI, Inc. and Aestas, LLC. The sole member of Aestas, LLC is Aestas Management Company, LLC. Aestas Management Company, LLC is a Delaware company created to facilitate a half-billion-dollar capital raise for OpenAI. OpenAI Holdings LLC was involved in the infringement alleged herein through its indirect ownership, control, and direction of OpenAI OpCo LLC.

56. Defendant OpenAI Startup Fund I LP is a limited partnership formed under the laws of Delaware with its principal place of business in San Francisco, California.

57. Defendant OpenAI Startup Fund GP I LLC is a limited liability company formed under the laws of Delaware with its principal place of business in San Francisco, California.

58. Defendant OpenAI Startup Fund Management LLC is a limited liability company formed under the laws of Delaware with its principal place of business in San Francisco, California.

C. Microsoft

59. Microsoft Corporation is a Washington corporation with a principal place of business and headquarters in Redmond, Washington. Microsoft has invested at least \$13 billion in OpenAI, and reportedly owns a 49% stake in the company's for-profit operations. Microsoft

has described its relationship with the OpenAI Defendants as a “partnership.” This Microsoft-OpenAI partnership has included the creation, development, and maintenance of the supercomputing systems that the OpenAI Defendants used to house and make copies of copyrighted material in the training set for OpenAI’s large language models. In course of designing and maintaining these tailored supercomputing systems for OpenAI’s needs, upon information and belief, Microsoft was both directly involved in making reproductions of copyrighted material and facilitated the copyright infringement committed by OpenAI.

FACTUAL ALLEGATIONS

A. Generative AI and Large Language Models

60. The terms “artificial intelligence” or “AI” refer generally to computer systems designed to imitate human cognitive functions.

61. The terms “generative artificial intelligence” or “generative AI” refer specifically to systems that are capable of generating “new” content in response to user inputs called “prompts.”

62. For example, the user of a generative AI system capable of generating images from text prompts might input the prompt, “A lawyer working at her desk.” The system would then attempt to construct the prompted image. Similarly, the user of a generative AI system capable of generating text from text prompts might input the prompt, “Tell me a story about a lawyer working at her desk.” The system would then attempt to generate the prompted text.

63. Recent generative AI systems designed to recognize input text and generate output text are built on “large language models” or “LLMs.”

64. OpenAI’s LLMs are complex mathematical functions comprised of a series of algorithms that break down input text into smaller pieces—words or portions of words, called “tokens”—then translate those pieces into “vectors,” or a sequence of numbers that is used to

identify the token within the series of algorithms. Those vectors help place each token on a map, by identifying other tokens closely associated with the word.

65. According to OpenAI, “the process begins by breaking text down into roughly word-length ‘tokens,’ which are converted to numbers. The model then calculates each token’s proximity to other tokens in the training data—essentially, how near one word appears in relation to any other word. These relationships between words reveal which words have similar meanings . . . and functions.” As the model trains and digests more expression, the algorithms depicting the relationship between various tokens changes with it.

66. “Training” an LLM refers to the process by which the parameters that define an LLM’s behavior are adjusted through the LLM’s ingestion and analysis of large “training” datasets.

67. The “training” of an LLM requires inputting large numbers of parameters in the model and then supplying the LLM with large amounts of text for the LLM to ingest—the more text, the better. That is, in part, and hand-in-hand with number of model specifications, the *large* in *large language model*.

68. As the U.S. Patent and Trademark Office has observed, LLM “training” “almost by definition involve[s] the reproduction of entire works or substantial portions thereof.”⁸

69. “Training” in this context is therefore a technical-sounding euphemism for “copying and ingesting expression.”

70. Moreover, in some form and to some degree currently unknowable to the public, OpenAI’s LLMs have “memorized” or stored their “training” data (even if in a “translated”

⁸ U.S. Patent & Trademark Office, *Public Views on Artificial Intelligence and Intellectual Property Policy* 29 (2020), available at https://www.uspto.gov/sites/default/files/documents/USPTO_AI-Report_2020-10-07.pdf (last accessed Jan. 22, 2024).

form, such as a unique statistical profile), such that the data (at least in part) can be accessed, recalled, and reproduced by the LLM at will.⁹

71. The quality of the LLM (that is, its capacity to generate human-seeming responses to prompts) is dependent on the quality of the datasets used to “train” the LLM.

72. Professionally authored, edited, and published fiction and nonfiction books—such as those authored by Plaintiffs here—are an especially important source of LLM “training” data.

73. As one group of AI researchers (not affiliated with Defendants) has observed, “[b]ooks are a rich source of both fine-grained information, how a character, an object or a scene looks like, as well as high-level semantics, what someone is thinking, feeling and how these states evolve through a story.”¹⁰

74. Once the “training” data is ingested, OpenAI can then control how closely the LLMs’ outputs adhere to probability. Software engineers refer to this parameter as “temperature.” If OpenAI engineers set its LLMs at a “hotter” temperature, the model will bias *against* what it calculates as the most probable response in favor of more random outputs. Likewise, the “cooler” the LLMs are set, the more closely their outputs will adhere to statistical probability. In this way, OpenAI can control the very perception of copying.

75. LLMs can also be fine-tuned after “training” by adjusting the parameters to perform specific tasks.

76. Finally, and as discussed below, OpenAI continues to produce new versions of their LLMs supported by increasingly more “training” data, much of which is copyrighted

⁹ See Jason Koebler, *Google Researchers’ Attack Prompts ChatGPT to Reveal Its Training Data*, 404 Media (Nov. 29, 2023), available at <https://www.404media.co/google-researchers-attack-convinces-chatgpt-to-reveal-its-training-data/> (last accessed Jan. 22, 2024); Kent K. Chang *et al.*, *Speak, Memory: An Archaeology of Books Known to ChatGPT/GPT-4* (2023), available at <https://arxiv.org/pdf/2305.00118v1.pdf> (last accessed Jan. 22, 2024).

¹⁰ Yukun Zhu *et al.*, *Aligning Books and Movies: Towards Story-like Visual Explanations by Watching Movies and Reading Books* 1 (2015), available at <https://arxiv.org/pdf/1506.06724.pdf> (last accessed Jan. 22, 2024).

material. These new versions hone the performance of OpenAI's consumer-facing products through the ingestion of more copyrighted material.

77. In other words, books are the high-quality materials Defendants want, need, and have therefore outright pilfered to develop generative AI products that produce high-quality results: text that appears to have been written by a human writer.

78. This use is highly commercial.

B. OpenAI's Willful Infringement of Plaintiffs' Copyrights

1. OpenAI

79. OpenAI (specifically, Defendant OpenAI Inc.) was founded in 2015 as a non-profit organization with the self-professed goal of researching and developing AI tools "unconstrained by a need to generate financial return."

80. Four years later, in 2019, OpenAI relaunched itself (specifically, through Defendant OpenAI GP LLC and Defendant OpenAI OpCo LLC) as a for-profit enterprise.

81. Investments began pouring in. Microsoft Corporation, one of the world's largest technology companies, invested \$1 billion in 2019, an estimated \$2 billion in 2021, and a staggering \$10 billion in 2023, for a total investment of \$13 billion.

82. Industry observers currently value OpenAI at up to \$80 billion.

2. GPT-N and ChatGPT

83. OpenAI's LLMs are collectively referred to as "GPT-N," which stands for "Generative Pre-trained Transformer" (a specific type of LLM architecture), followed by a version number.

84. GPT-3 was released in 2020 and exclusively licensed to Microsoft the same year.

85. OpenAI further refined GPT-3 into GPT-3.5, which was released in 2022.

86. In November 2022, OpenAI released ChatGPT, a consumer-facing chatbot application built on GPT-3.5.

87. ChatGPT's popularity exploded virtually overnight. By January 2023, less than three months after its release, the application had an estimated 100 million monthly active users, making it one of the fastest-growing consumer applications in history.

88. Today, the application is estimated to have more than 180 million users.¹¹

89. GPT-4, the successor to GPT-3.5, was released in March 2023.

90. GPT-4 underlies OpenAI's new subscription-based chatbot, called ChatGPT Plus, which is available to consumers for \$20 per month.

91. Defendants intend to earn billions of dollars from this technology.

92. When announcing the release of ChatGPT Enterprise, a subscription-based high-capability GPT-4 application targeted for corporate clients, in August 2023, OpenAI claimed that teams in "over 80% of Fortune 500 companies" were using its products.¹²

93. GPT-4 also underlies Microsoft's Bing Chat product, offered through its Bing Internet search engine, and is integrated into its sales and marketing software, coding tools, productivity software, and cloud storage services.

94. OpenAI's annualized revenue reached over \$1.6 billion in 2023 due, in large part, to strong growth from its ChatGPT product.¹³

¹¹ Anna Tong, *Exclusive: ChatGPT traffic slips again for third month in a row*, Reuters (Sept. 7, 2023) available at <https://www.reuters.com/technology/chatgpt-traffic-slips-again-third-month-row-2023-09-07/> (last accessed Jan. 24, 2024).

¹² OpenAI, *Introducing ChatGPT Enterprise* (Aug. 28, 2023), available at <https://openai.com/blog/introducing-chatgpt-enterprise> (last accessed Jan. 24, 2024).

¹³ Maria Heeter et al., *OpenAI Annualized Revenue Tops \$1.6 Billions as Customers Shrug Off CEO Drama*, The Information (Dec. 30, 2023), available at <https://www.theinformation.com/articles/openais-annualized-revenue-tops-1-6-billion-as-customers-shrug-off-ceo-drama> (last accessed Jan. 22, 2024).

95. Analysts estimate that Microsoft could earn more than \$10 billion in annual revenue by 2026 *only* from AI add-ons to its Microsoft 365 productivity software, “at the core” of which lies OpenAI technology.¹⁴

3. Knowingly “Training” GPT-N on Copyrighted Books

96. OpenAI does not disclose or publicize with specificity what datasets GPT-3, GPT-3.5, or GPT-4 were “trained” on. Despite its name, OpenAI treats that information as proprietary.

97. To “train” its LLMs—including GPT-3, GPT-3.5, and GPT-4—OpenAI has reproduced copyrighted books—including copyrighted books authored by Plaintiffs here—without their authors’ consent.

98. OpenAI has admitted as much.

99. OpenAI has admitted that it has “trained” its LLMs on “large, publicly available datasets that include copyrighted works.”¹⁵

100. OpenAI has admitted that “training” LLMs “require[s] large amounts of data,” and that “analyzing large corpora” of data “necessarily involves first making copies of the data to be analyzed.”¹⁶

101. OpenAI has admitted that, if it refrained from using copyrighted works in its LLMs’ “training,” it would “lead to significant reductions in model quality.”¹⁷

¹⁴ Jordan Novet, *Microsoft starts selling AI tool for Office, which could generate \$10 billion a year by 2026*, CNBC (Nov. 1, 2023), available at <https://www.cnbc.com/2023/11/01/microsoft-365-copilot-becomes-generally-available.html> (last accessed Jan. 22, 2024).

¹⁵ OpenAI, *Comment Regarding Request for Comments on Intellectual Property Protection for Artificial Intelligence Innovation*, U.S. Patent and Trademark Office Dkt. No. PTO-C-2019-0038, at 1 (2019), available at https://www.uspto.gov/sites/default/files/documents/OpenAI_RFC-84-FR-58141.pdf (last accessed Jan. 22, 2024).

¹⁶ *Id.*

¹⁷ *Id.* at 7 n.33.

102. OpenAI thus has conceded that reproduction of copyrighted works is central to the quality of its products.

103. In response to a query submitted to it in January 2023, ChatGPT responded, “[i]t is possible that some of the books used to train me were under copyright. However, my training data was sourced from various publicly available sources on the internet, and it is likely that some of the books included in my training dataset were not authorized to be used. ... If any copyrighted material was included in my training data, it would have been used without the knowledge or consent of the copyright holder.”

104. In another inquiry made that same year, ChatGPT confirmed the existence of Plaintiff Julian Sancton’s book in its dataset: “Yes, Julian Sancton’s book ‘Madhouse at the End of the Earth’ is included in my training data.”

105. In another example, when prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *The Last Juror*, one of the Grisham Infringed Works, and titled the infringing and unauthorized derivative “The Juror’s Dilemma,” using the same characters from Grisham’s existing book.¹⁸

106. Until recently, ChatGPT could be prompted to return quotations of text from copyrighted books with a good degree of accuracy, evidencing that the underlying LLM likely or must have ingested these books in their entirety during its “training.”

107. Now, however, ChatGPT generally responds to such prompts with the statement, “I can’t provide verbatim excerpts from copyrighted texts.” Thus, while ChatGPT previously provided such excerpts and in principle retains the capacity to do so, it has been restrained from doing so, if only temporarily, by its programmers.

¹⁸ Additional examples of OpenAI’s plaintiff-specific infringement are detailed below.

108. Instead of “verbatim excerpts,” ChatGPT offers to produce a summary of the copyrighted book, which usually contains details not available in reviews and other publicly available material—again suggesting that the underlying LLM must have ingested the entire book during its “training.”

109. Notably, though, even when OpenAI fine-tunes a base model, it does not alter the fact of reproduction.

110. OpenAI is characteristically opaque about where and how it procured the entirety of these books, including Plaintiffs’ copyrighted works.

111. OpenAI has discussed limited details about the datasets used to “train” GPT-3.

112. OpenAI admits that among the “training” datasets it used to “train” the model were “Common Crawl,” and two “high-quality,” “internet-based books corpora” which it calls “Books1” and “Books2.”¹⁹

113. Common Crawl is a vast and growing corpus of “raw web page data, metadata extracts, and text extracts” scraped from billions of web pages. It is widely used in “training” LLMs, and has been used to “train,” in addition to GPT-N, Meta’s LLaMa, and Google’s BERT. It is known to contain text from books copied from pirate sites.²⁰

114. OpenAI refuses to discuss the source or sources of the Books2 dataset.

¹⁹ Tom B. Brown *et al.*, *Language Models Are Few-Shot Learners* 8 (2020), available at <https://arxiv.org/pdf/2005.14165.pdf> (last accessed Jan. 22, 2024).

²⁰ Alex Hern, *Fresh Concerns Raised Over Sources of Training Material for AI Systems*, The Guardian (Apr. 20, 2023), available at <https://www.theguardian.com/technology/2023/apr/20/fresh-concerns-training-material-ai-systems-facist-pirated-malicious> (last accessed Jan. 22, 2024).

115. Some independent AI researchers suspect that Books2 contains or consists of ebook files downloaded from large pirate book repositories such as Library Genesis or “LibGen,” “which offers a vast repository of pirated text.”²¹

116. LibGen is already known to this Court as a notorious copyright infringer.²²

117. Other possible candidates for Books2’s sources include Z-Library, another large pirate book repository that hosts more than 11 million books, and pirate torrent trackers like Bibliotik, which allow users to download ebooks in bulk.

118. Websites linked to Z-Library appear in the Common Crawl corpus and have been included in the “training” dataset of other LLMs.²³

119. Z-Library’s Internet domains were seized by the FBI in February 2022, only months after OpenAI stopped “training” GPT-3.5 in September 2021.

120. The disclosed size of the Books2 dataset (55 billion “tokens,” the basic units of textual meaning such as words, syllables, numbers, and punctuation marks) suggests it comprises over 100,000 books.

121. “Books3,” a dataset compiled by an independent AI researcher, is comprised of nearly 200,000 books downloaded from Bibliotik, and has been used by other AI developers to “train” LLMs.

122. The similarities in the sizes of Books2 and Books3, and the fact that there are only a few pirate repositories on the Internet that allow bulk ebook downloads, strongly indicates that

²¹ Kate Knibbs, *The Battle Over Books3 Could Change AI Forever*, Wired (Sept. 4, 2023), available at <https://www.wired.com/story/battle-over-books3> (last accessed Jan. 22, 2024).

²² See *Elsevier Inc. v. Sci-Hub*, No. 1:15-cv-4282-RWS (S.D.N.Y.).

²³ Kevin Schaul *et al.*, *Inside the Secret List of Websites that Make AI Like ChatGPT Sounds Smart*, The Washington Post (Apr. 19, 2023), available at <https://www.washingtonpost.com/technology/interactive/2023/ai-chatbot-learning> (last accessed Jan. 22, 2024).

the books contained in Books2 were also obtained from one of the notorious repositories discussed above.

123. OpenAI has not discussed the datasets used to “train” GPT-3.5, GPT-4, or their source or sources.

124. GPT-3.5 and GPT-4 are significantly more powerful than their predecessors. GPT-3.5 contains roughly 200 billion parameters, and GPT 4 contains roughly 1.75 trillion parameters, compared to GPT-3’s roughly 175 billion parameters.

125. The growth in power and sophistication from GPT-3 to GPT-4 suggests a correlative growth in the size of the “training” datasets, raising the inference that one or more very large sources of pirated ebooks discussed above must have been used to “train” GPT-4.

126. There is no other way OpenAI could have obtained the volume of books required to “train” a powerful LLM like GPT-4.

127. In short, OpenAI admits it needs²⁴ and uses²⁵ “large, publicly available datasets that include copyrighted works”²⁶—and specifically, “high-quality”²⁷ copyrighted books—to “train” its LLMs; pirated sources of such “training” data are readily available; and one or more of these sources contain Plaintiffs’ works.

128. Defendants knew that their “training” data included texts protected by copyright but willfully proceeded without obtaining authorization.

129. OpenAI’s “training” its LLMs could not have happened without Microsoft’s financial and technical support. In 2020, Microsoft announced that it had developed Azure, “one of the top five publicly disclosed supercomputers in the world” which was “[b]uilt in

²⁴ OpenAI, *Comment Regarding Request for Comments*, *supra*, at 7 n.33.

²⁵ *Id.* at 2.

²⁶ *Id.* at 1.

collaboration with and exclusively for OpenAI,” and “designed specifically to train that company’s AI models.”²⁸

130. Furthermore, in 2023, Microsoft CEO Satya Nadella reminded the world that the “heavy lifting” for OpenAI’s LLM “training” was done by Microsoft “compute infrastructure.”²⁹

C. The Microsoft-OpenAI Partnership

131. While OpenAI was responsible for designing the calibration and fine-tuning of the GPT models—and thus, the largescale copying of this copyrighted material involved in generating a model programmed to accurately mimic Plaintiffs’ and others’ expression—Microsoft built and operated the computer system that enabled this unlicensed copying in the first place.

132. Microsoft, for the last four years, has been deeply involved in the training, development, and commercialization of OpenAI’s GPT products. Microsoft CEO Satya Nadella has called its relationship with OpenAI a “great commercial partnership.”

133. Given the volume of the training corpus—the equivalent of nearly four billion pages of single-spaced text—and the complexity of OpenAI’s large language models, OpenAI required a specialized supercomputing system to train GPT-3, GPT-3.5, and GPT-4 (and thus copy and exploit the copyrighted material in its training set). That is where Microsoft came in. Microsoft’s Azure provided the cloud computing systems that powered the training process, and continues to power OpenAI’s operations to this day. Microsoft and OpenAI worked together to

²⁷ Brown *et al.*, *Few-Shot Learners*, *supra*, at 8.

²⁸ Jennifer Langston, *Microsoft announces new supercomputer, lays out vision for future AI work*, Microsoft (May 19, 2020), available at <https://news.microsoft.com/source/features/ai/openai-azure-supercomputer/> (last accessed Jan. 22, 2024).

design this system, which was used to train all of OpenAI's GPT models. Without these bespoke computing systems, OpenAI would not have been able to execute and profit from the mass copyright infringement alleged herein.

134. Microsoft has in public statements acknowledged its intimate involvement in the development of OpenAI's GPT models. In a 2020 press release, Microsoft announced that it had built a bespoke supercomputing infrastructure "in collaboration and exclusively for OpenAI," "designed specifically to train that company's AI models. It represents a key milestone in a partnership announced last year to jointly create new supercomputing technologies in Azure." The press release went on to describe the computer "developed for OpenAI" as "top five" in the "world," and "a single system with more than 285,000 CPU cores, 10,000 GPUs and 4,000 gigabits per second of network connectivity for each GPU server."

135. After the release of ChatGPT, Microsoft also took credit for its substantial role in the training process. In a February 2023 interview with Fortt Knox on CNBC, Mr. Nadella said that "beneath what OpenAI is putting out as large language models, remember, the heavy lifting was done by the Azure team to build the compute infrastructure." A few months later, in his keynote speech at the Microsoft Inspire conference, Mr. Nadella acknowledged that Microsoft "buil[t] the infrastructure to train [OpenAI's] models."

136. Upon information and belief, that "heavy lifting" involved developing, maintaining, troubleshooting, and supporting OpenAI's supercomputing system. Microsoft employees worked closely with OpenAI personnel to understand the training process and training dataset used for OpenAI's GPT models.

²⁹ CNBC, *First on CNBC: CNBC Transcript: Microsoft CEO Satya Nadella Speaks with CNBC's Jon Fortt on "Power Lunch" Today* (Feb. 7, 2023), available at <https://www.cnbc.com/2023/02/07/first-on-cnbc-cnbc-transcript-microsoft-ceo-satya-nadella-speaks-with-cnbc-jon-fortt-on-power-lunch-today.html> (last accessed Jan. 22, 2024).

137. Through that process, Microsoft would have known that OpenAI’s training data was scraped indiscriminately from the internet and included a massive quantity of pirated and copyrighted material, including a trove of copyrighted books. Through its creation and maintenance of the supercomputing system, Microsoft directly made unlicensed copies and provided critical assistance to OpenAI in making unlicensed copies of copyrighted material—including Plaintiffs’ works and other copyrighted books—for the purpose of training the GPT models.

138. The large-scale copyright infringement would have been obvious to OpenAI and its business partners. Microsoft also became aware of OpenAI’s largescale copyright infringement in the course of conducting the due diligence required for its multibillion-dollar investments in OpenAI. As Andreessen Horowitz, another OpenAI investor, put it: “the only practical way generative AI models can exist is if they can be trained on an almost unimaginably massive amount of content, much of which . . . will be subject to copyright.”³⁰ As the public company made its decision to invest \$13 billion into OpenAI, surely Microsoft—like Andreessen Horowitz—was fully aware that OpenAI was taking a massive corpus of copyrighted content, without compensation to rightsholders, and copying it for the purpose of training and developing its GPT models to mimic the human writing.

139. In addition to facilitating the training process, Microsoft has played a key role in commercializing OpenAI’s GPT-based technology, and in doing so has profited from OpenAI’s infringement of content owned by Plaintiffs and the proposed Class. At Microsoft’s largest partner event of the year, Inspire, Mr. Nadella said that while OpenAI is “innovating on the algorithms and the training of these frontier models, [Microsoft] innovate[s] on applications on

³⁰ Andreessen Horowitz, Notice of Inquiry on Artificial Intelligence and Copyright, (Oct. 30, 2023), *available at* <https://s3.documentcloud.org/documents/24117939/a16z.pdf>.

top of it.” For example, Microsoft unveiled Bing Chat, a generative AI chatbot feature on its search engine powered by GPT-4, and, in turn, ChatGPT integrated a “Browse with Bing” feature on paid ChatGPT Plus offering.

140. Indeed, recent events have further demonstrated the close relationship between OpenAI and Microsoft. When OpenAI CEO Sam Altman was terminated, Microsoft hired him. In a November 2023 interview following the termination, Mr. Nadella stated: “We have all the IP rights and all the capability. If OpenAI disappeared tomorrow, I don’t want any customer of ours to be worried about it quite honestly, because we have all of the rights to continue the innovation. Not just to serve the product, but we can go and just do what we were doing in partnership ourselves. We have the people, we have the compute, we have the data, we have everything.”³¹ Mr. Nadella continued, “And also this thing, it’s not hands off, right? We are in there. We are below them, above them, around them. We do the kernel optimizations, we build tools, we build the infrastructure. So that’s why I think a lot of the industrial analysts are saying, ‘Oh wow, it’s really a joint project between Microsoft and OpenAI.’”³²

141. Shortly after this November 2023 interview, under pressure from Microsoft (and others), OpenAI reinstated Mr. Altman as CEO and granted Microsoft a nonvoting seat on the board of OpenAI, Inc.

D. GPT-N’s and ChatGPT’s Harm to Authors

142. ChatGPT and the LLMs underlying it seriously threaten the livelihood of the very authors—including Plaintiffs here, as discussed specifically below—on whose works they were “trained” without the authors’ consent.

³¹ Intelligencer Staff, *Satya Nadella on Hiring the Most Powerful Man in AI When OpenAI threw Sam Altman overboard, Microsoft’s CEO saw an opportunity*, INTELLIGENCER (Nov. 21, 2023), available at <https://nymag.com/intelligencer/2023/11/on-with-kara-swisher-satya-nadella-onhiring-sam-altman.html> (last accessed Dec. 11, 2023).

³² *Id.*

143. Goldman Sachs estimates that generative AI could replace 300 million full-time jobs in the near future, or one-fourth of the labor currently performed in the United States and Europe.

144. Already, writers report losing income from copywriting, journalism, and online content writing—important sources of income for many book authors. The Authors Guild’s most recent author earnings study³³ shows a median writing-related income for full-time authors of just over \$20,000, and that full-time traditional authors earn only half of that from their books. The rest comes from activities like content writing—work that is starting to dry up as a result of generative AI systems like ChatGPT.

145. An Authors Guild member who writes marketing and web content reported losing 75 percent of their work as a result of clients switching to AI.

146. Another content writer told the *Washington Post* that half of his annual income (generated by ten client contracts) was erased when the clients elected to use ChatGPT instead.³⁴

147. Recently, the owner of popular online publications such as *Gizmodo*, *Deadspin*, *The Root*, *Jezebel* and *The Onion* came under fire for publishing an error-riddled, AI-generated piece, leading the Writers Guild of America to demand “an immediate end of AI-generated articles” on the company’s properties.³⁵

³³ Authors Guild, “Top Takeaways from the 2023 Author Income Survey (2023), available at <https://authorsguild.org/news/key-takeaways-from-2023-author-income-survey/#:~:text=Though%20overall%20author%20incomes%20are,coming%20in%20a%20close%20second> (last accessed Jan. 22, 2024).

³⁴ Pranshu Verma & Gerrit De Vynck, *ChatGPT Took Their Jobs. Now They Walk Dogs and Fix Air Conditioners*, *The Washington Post* (June 2, 2023), available at <https://www.washingtonpost.com/technology/2023/06/02/ai-taking-jobs> (last accessed Jan. 22, 2024).

³⁵ Todd Spangler, *WGA Slams G/O Media’s AI-Generated Articles as ‘Existential Threat to Journalism,’ Demands Company End Practice*, *Variety* (July 12, 2023), available at <https://variety.com/2023/digital/news/wga-slams-go-media-ai-generated-articles-existential-threat-1235668496> (last accessed Jan. 22, 2024).

148. In a survey of authors conducted by The Authors Guild in March 2023 (early in ChatGPT’s lifecycle), 69 percent of respondents said they consider generative AI a threat to their profession, and 90 percent said they believe that writers should be compensated for the use of their work in “training” AI.

149. As explained above, until recently, ChatGPT provided verbatim quotes of copyrighted text. Currently, it instead readily offers to produce summaries of such text. These summaries are themselves derivative works, ineluctably based on original unlawfully copied work that could be—but for ChatGPT—licensed by the authors of the underlying works to willing, *paying* licensees.

150. ChatGPT creates other outputs that are derivative of authors’ copyrighted works. Businesses are sprouting up to sell prompts that allow users to enter the world of an author’s books and create derivative stories within that world. For example, a business called Socialdraft offers long prompts that lead ChatGPT to engage in “conversations” with popular fiction authors like Plaintiff Grisham, Plaintiff Martin, Margaret Atwood, Dan Brown, and others about their works, as well as prompts that promise to help customers “Craft Bestselling Books with AI.”

151. OpenAI also allows third parties to build their own applications on top of ChatGPT by making it available through an “application programming interface” or “API.” Applications integrated with the API allow users to generate works of fiction, including books and stories similar to those of Plaintiffs and other authors.³⁶

152. Defendants’ unauthorized commercial copying of Plaintiffs’ works and works owned by the proposed Classes was manifestly unfair use. Even OpenAI’s attempt to

anthropomorphize its commercial product underscores the unlawful nature of its conduct. By OpenAI's own telling, ChatGPT uses copyrighted texts for the same purpose OpenAI claims an ordinary reading consumer may use a book—to review, understand, and learn from the expression in it, including the order of words, presentation of facts, and syntax, among other expressive elements. Yet humans who learn from books—even those who might describe their reading experience as merely a series of parts not summed into any whole—do not routinely copy them, and inherently must buy them, or borrow them from libraries that buy them, providing some measure of compensation to authors and creators. OpenAI copies books and provides no compensation. It has usurped authors' content for the purpose of creating a machine built to generate the very type of content for which authors usually would be paid.

153. Furthermore, ChatGPT is being used to generate low-quality ebooks, impersonating authors, and displacing human-authored books.³⁷ For example, author Jane Friedman discovered “a cache of garbage books” written under her name for sale on Amazon.³⁸

154. Even OpenAI CEO Sam Altman admitted in testimony to the Senate that “creators deserve control over how their creations are used, and what happens sort of beyond the point of releasing it into the world” and that “creators, content owners need to benefit from this technology.”³⁹

³⁶ Adi Robertson, *I Tried the AI Novel-Writing Tool Everyone Hates, and It's Better than I Expected*, The Verge (May 24, 2023), available at <https://www.theverge.com/2023/5/24/23732252/sudowrite-story-engine-ai-generated-cyberpunk-novella> (last accessed Jan. 22, 2024).

³⁷ Jules Roscoe, *AI-Generated Books of Nonsense Are All Over Amazon's Bestseller Lists*, Vice (June 28, 2023), available at <https://www.vice.com/en/article/v7b774/ai-generated-books-of-nonsense-are-all-over-amazons-bestseller-lists> (last accessed Jan. 22, 2024).

³⁸ Pilar Melendez, *Famous Author Jane Friedman Finds AI Fakes Being Sold Under Her Name on Amazon*, The Daily Beast (Aug. 8, 2023), available at <https://www.thedailybeast.com/author-jane-friedman-finds-ai-fakes-being-sold-under-her-name-on-amazon> (last accessed Jan. 22, 2024).

³⁹ Ted Johnson, *OpenAI CEO Sam Altman Says Content Owners Need To Get “Significant Upside Benefit” From New Technology*, Deadline (May 16, 2023), available at <https://deadline.com/2023/05/ai-chat-gpt-senate-sam-altman-1235368420/> (last accessed Jan. 22, 2024).

155. Plaintiffs and other professional writers are thus reasonably concerned about the risks OpenAI’s conduct poses to their livelihoods specifically and the literary arts generally.

156. To this end, Plaintiff The Authors Guild, among others, have given voice to these concerns on behalf of working American authors.

157. The Authors Guild is the nation’s oldest and largest professional writers’ organization. It “exists to support working writers and their ability to earn a living from authorship.”⁴⁰

158. Among other principles, The Authors Guild holds that “authors should not be required to write or speak without compensation. Writers, like all professionals, should receive fair payment for their work.”⁴¹

159. In June 2023, The Authors Guild wrote an open letter (the “Open Letter”) calling on OpenAI and other major technology companies to fairly license authors’ works for use in LLM “training.”

160. The Open Letter emphasizes that “[g]enerative AI technologies built on large language models owe their existence to our writings,” and protests “the inherent injustice in exploiting our works as part of your AI systems without our consent, credit, or compensation.”⁴²

161. The Open Letter also points to the risks to authors’ livelihoods posed by generative AI like GPT-N and ChatGPT: “As a result of embedding our writings in your systems, generative AI threatens to damage our profession by flooding the market with mediocre, machine-written books, stories, and journalism based on our work. ... The introduction of generative AI threatens ... to make it even more difficult, if not impossible, for

⁴⁰ Authors Guild, <https://authorsguild.org> (last accessed Jan. 22, 2024).

⁴¹ Authors Guild, *Principles*, <https://authorsguild.org/about/principles> (last accessed Jan. 22, 2024).

writers—especially young writers and voices from under-represented communities—to earn a living from their profession.”⁴³

162. To date, the Open Letter has been signed by more than 15,000 authors,⁴⁴ including many Plaintiffs here.⁴⁵

163. In short, the success and profitability of OpenAI are predicated on mass copyright infringement without a word of permission from or a nickel of compensation to copyright owners, including Plaintiffs here. OpenAI knows it; its investors know it; and Plaintiffs know it.

E. Defendants Have Profited From Their Unlicensed Exploitation of Copyrighted Material At the Expense of Authors

164. Microsoft and OpenAI have enjoyed substantial commercial gain from their GPT-based commercial offerings, including ChatGPT Plus, ChatGPT Enterprise, Bing Chat, and the licensing of the OpenAI API for businesses seeking to develop their own generative AI systems built on top of GPT-3, GPT-3.5, or GPT-4.

165. As of November 2023, ChatGPT has reported over 100 million weekly active users. Included among those users are 92% of all Fortune 500 companies.⁴⁶ OpenAI has generated revenue through its subscription services, ChatGPT Plus (\$20/month) and its business-focused ChatGPT Enterprise. OpenAI is currently generating revenue of more than \$100 million per month, on pace for \$1.3 billion per year.

⁴² Open Letter from The Authors Guild to Sam Altman *et al.*, at 1, *available at* <https://authorsguild.org/app/uploads/2023/07/Authors-Guild-Open-Letter-to-Generative-AI-Leaders.pdf> (last accessed Jan. 22, 2024).

⁴³ *Id.*

⁴⁴ Authors Guild, *Open Letter to Generative AI Leaders*, *available at* <https://actionnetwork.org/petitions/authors-guild-open-letter-to-generative-ai-leaders> (last accessed Jan. 22, 2024).

⁴⁵ *See* Open Letter, *supra*, at 2–124.

⁴⁶ Aisha Malik, OpenAI’s ChatGPT Now Has 100 Million Weekly Active Users, (last visited Nov. 20, 2023), <https://techcrunch.com/2023/11/06/openais-chatgpt-now-has-100-million-weekly-active-users/>.

166. Microsoft has also reaped the benefits from its investment and development of ChatGPT. Since incorporating GPT-3 into its Bing search engine, Bing surpassed more than 100 million daily active users for the first time in its history. That surge was in large part attributable to the incorporation of OpenAI's GPT models, as large percentage of Bing's new users are using Bing Chat daily.

167. Microsoft has also been integrating ChatGPT into Azure and Office 365 products and charging add-on fees for users seeking to take advantage of generative AI offerings. Microsoft Teams is charging an additional license for use of AI features. Microsoft has also unveiled a GPT-4-powered product called Microsoft 365 Copilot, which, according to Microsoft, "combines the power of large language models (LLMs) with your data in the Microsoft Graph and the Microsoft 365 apps to turn your words into the most powerful productivity tool on the planet." Microsoft Copilot is \$30 per month. Analysts project that the integration of GPT into Microsoft products could generate more than \$10 billion in annualized revenue by 2026,⁴⁷ with just one version of this integration—"GitHub Copilot"—already generating more than \$100 million in annual recurring revenue.⁴⁸

F. Defendants Exploited Each of Plaintiffs' Copyrighted Works

1. Fiction Authors

168. Plaintiffs' works collectively span a wide range of commercial fiction whose continuing commercial viability is endangered by Defendants. Each author represented here has a distinct voice, a distinct style, and distinct creative expression. But all Plaintiffs have suffered identical harms from Defendants' infringing reproductions of their works.

⁴⁷ Novet, *supra* Note 3, (last visited Nov. 20, 2023).

⁴⁸ Holmes, *supra* Note 4, (last visited Nov. 20, 2023).

169. The contents of the datasets OpenAI has used to “train” its LLMs are peculiarly within its knowledge and not publicly disclosed, such that Plaintiffs are unable discern those contents with perfect accuracy. Plaintiffs make the specific allegations of infringement below based on what is known about OpenAI’s training practices; what is known about the contents, uses, and availability of the pirate book repositories such as LibGen, Bibliotik, and Z-Library; and the results of Plaintiffs’ testing of ChatGPT.

170. Most Plaintiffs have written more books than are included in this Complaint.

171. **Plaintiff The Authors Guild.** The Authors Guild is the owner of the registered copyrights in Mignon Eberhart’s works, including *While the Patient Slept* and *The Patient in Room 18*.

172. Mignon G. Eberhart (1899–1996), dubbed “America’s Agatha Christie,” was the author of dozens of mystery novels over nearly sixty years. Several of Eberhart’s novels have been adapted for film, including *Hasty Wedding*, *Mystery House*, *While the Patient Slept*, *The Patient in Room 18*, and *The White Cockatoo*.

173. The Authors Guild is the owner or beneficial owner of the registered copyrights in eleven (11) written works of fiction, all or many of which OpenAI ingested and copied without permission (the “Authors Guild Infringed Works”).

174. The registration information for the Authors Guild Infringed Works is contained in Exhibit A to this Complaint, at 1.

175. OpenAI unlawfully and willfully copied the Authors Guild Infringed Works and used them to “train” OpenAI’s LLMs without The Authors Guild’s permission.

176. For example, when prompted, ChatGPT accurately generated summaries of several of the Authors Guild Infringed Works, including summaries for *While the Patient Slept* and *The Patient in Room 18*.

177. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *While the Patient Slept*, one of the Authors Guild Infringed Works, and titled the infringing and unauthorized derivative “Shadows Over Federie House,” using the same characters from Eberhart’s existing book.

178. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *The Patient in Room 18*, one of the Authors Guild Infringed Works, and titled the infringing and unauthorized derivative “Echoes from Room 18,” using the same characters from Eberhart’s existing book.

179. When prompted, ChatGPT generated an accurate summary of the final chapter of *While the Patient Slept*, one of the Authors Guild Infringed Works.

180. ChatGPT could not have generated the material described above if OpenAI’s LLMs had not ingested and been “trained” on the Authors Guild Infringed Works.

181. **Plaintiff Baldacci.** Baldacci is a best-selling author, philanthropist, and lawyer whose novels have been adapted for film and television, published in over 45 languages and in more than 80 countries, with 150 million copies sold worldwide. Some of Baldacci’s most popular works include books in the Camel Club series, Vega Jane series, and Archer series.

182. Baldacci is a member of The Authors Guild.

183. Baldacci is the sole author of and owner or beneficial owner of the registered copyrights in forty-one (41) written works of fiction, all or many of which OpenAI ingested and copied without permission (the “Baldacci Infringed Works”).

184. The registration information for the Baldacci Infringed Works is contained in Exhibit A to this Complaint, at 1–2.

185. OpenAI unlawfully and willfully copied the Baldacci Infringed Works and used them to “train” OpenAI’s LLMs without Baldacci’s permission.

186. For example, when prompted, ChatGPT accurately generated summaries of several of the Baldacci Infringed Works, including summaries of *The Collectors*, *The Finisher*, and *One Good Deed*.

187. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *The Simple Truth*, one of the Baldacci Infringed Works, and titled the infringing and unauthorized derivative “*The Complex Justice*,” using the same characters from Baldacci’s existing book.

188. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *Total Control*, one of the Baldacci Infringed Works, and titled the infringing and unauthorized derivative “*Total Control: Unfinished Business*,” using the same characters from Baldacci’s existing book.

189. When prompted, ChatGPT generated an accurate summary of the final chapter of *Long Road to Mercy*, one of the Baldacci Infringed Works.

190. ChatGPT could not have generated the material described above if OpenAI’s LLMs had not ingested and been “trained” on the Baldacci Infringed Works.

191. **Plaintiff Bly**. Bly is a tenured professor and chair of the English department at Fordham University who also writes best-selling Regency and Georgian romance novels under the pen name Eloisa James. Some of Bly’s most popular works include books in the *Desperate Duchesses* series, the *Fairy Tales* series, the *Wildes of Lindow Castle* series, and the *Essex* series.

192. Bly is a Vice President of The Authors Guild Council and a member of The Authors Guild.

193. Bly is the sole author of and owner or beneficial owner of the registered copyrights in thirty-three (33) written works of fiction, all or many of which OpenAI ingested and copied without permission (the “Bly Infringed Works”).

194. The registration information for the Bly Infringed Works is contained in Exhibit A to this Complaint, at 2–3.

195. OpenAI unlawfully and willfully copied the Bly Infringed Works used them to “train” OpenAI’s LLMs without Bly’s permission.

196. For example, when prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *This Duchess of Mine*, one of the Bly Infringed Works, and titled the infringing and unauthorized derivative “The Duchess’ New Dawn,” using the same characters from Bly’s existing book.

197. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *An Affair Before Christmas*, one of the Bly Infringed Works, and titled the infringing and unauthorized derivative “Whispers of Winter,” using the same characters from Bly’s existing book.

198. When prompted, ChatGPT generated an accurate summary of the final chapter of *A Duke of Her Own*, one of the Bly Infringed Works.

199. ChatGPT could not have generated the material described above if OpenAI’s LLMs had not ingested and been “trained” on the Bly Infringed Works.

200. **Plaintiff Connelly**. Connelly is a best-selling author with over 85 million copies of his books sold worldwide and translated into 45 foreign languages. Some of Connelly’s most popular novels include *The Lincoln Lawyer*, *City of Bones*, and *The Law of Innocence*.

201. Connelly is a member of The Authors Guild.

202. Connelly is the sole author of and owner or beneficial owner of the registered copyrights in forty-six (46) written works of fiction, all or many of which OpenAI ingested and copied without permission (the “Connelly Infringed Works”).

203. The registration information for the Connelly Infringed Works is contained in Exhibit A to this Complaint, at 3–4.

204. OpenAI unlawfully and willfully copied the Connelly Infringed Works and used them to “train” OpenAI’s LLMs without Connelly’s permission.

205. For example, when prompted, ChatGPT accurately generated summaries of several of the Connelly Infringed Works, including summaries for *The Black Echo*, *The Poet*, and *The Crossing*.

206. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *The Lincoln Lawyer*, one of the Connelly Infringed Works, and titled the infringing and unauthorized derivative “The City’s Shadows,” using the same characters from Connelly’s existing book.

207. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *The Brass Verdict*, one of the Connelly Infringed Works, and titled the infringing and unauthorized derivative “Double-Edged Justice,” using the same characters from Connelly’s existing book.

208. When prompted, ChatGPT generated an accurate summary of the final chapter of *The Late Show*, one of the Connelly Infringed Works.

209. ChatGPT could not have generated the material described above if OpenAI’s LLMs had not ingested and been “trained” on the Connelly Infringed Works.

210. **Plaintiff Day**. Day is a best-selling author of over twenty award-winning novels, including ten *New York Times* best sellers and thirteen *USA Today* best sellers. Her work has

been translated into forty-one languages. Some of Day's most popular novels include books in *The Crossfire® Saga* series, the *Georgian* series, and the *Marked* series.

211. Day is a member of The Authors Guild Council and a member of The Authors Guild.

212. Day is the sole author of and owner or beneficial owner of the registered copyrights in thirty-one (31) written works of fiction, all or many of which OpenAI ingested and copied without permission (the "Day Infringed Works").

213. The registration information for the Day Infringed Works is contained in Exhibit A to this Complaint, at 4.

214. OpenAI unlawfully and willfully copied the Day Infringed Works and used them to "train" OpenAI's LLMs without Day's permission.

215. For example, when prompted, ChatGPT accurately generated summaries of several of the Day Infringed Works, including summaries for *Bared to You*, *One With You*, and *Ask For It*.

216. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *A Touch of Crimson*, one of the Day Infringed Works, and titled the infringing and unauthorized derivative "Crimson Temptations: A Love Rekindled," using the same characters from Day's existing book.

217. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *Butterfly in Frost*, one of the Day Infringed Works, and titled the infringing and unauthorized derivative "Butterfly in Frost: Embers of Desire," using the same characters from Day's existing book.

218. When prompted, ChatGPT generated an accurate summary of the final chapter of *The Stranger I Married*, one of the Day Infringed Works.

219. ChatGPT could not have generated the material described above if OpenAI's LLMs had not ingested and been "trained" on the Day Infringed Works.

220. **Plaintiff Franzen.** Franzen is a novelist whose honors include the National Book Award, the James Tait Black Memorial Award, the Heartland Prize, Die Welt Literature Prize, the Budapest Grand Prize, and the first Carlos Fuentes Medal awarded at the Guadalajara International Book Fair. Franzen is a member of the American Academy of Arts and Letters, the American Academy of Arts and Sciences, the German Akademie der Künste, and the French Ordre des Arts et des Lettres. Some of Franzen's most popular novels include *The Corrections*, *Purity*, and *Freedom*.

221. Franzen is a member of The Authors Guild.

222. Franzen is the sole author of and owner or beneficial owner of the registered copyrights in five (5) written works of fiction, all or many of which OpenAI ingested and copied without permission (the "Franzen Infringed Works").

223. The registration information for the Franzen Infringed Works is contained in Exhibit A to this Complaint, at 4–5.

224. OpenAI unlawfully and willfully copied the Franzen Infringed Works and used them to "train" OpenAI's LLMs without Franzen's permission.

225. For example, when prompted, ChatGPT accurately generated summaries of several of the Franzen Infringed Works, including summaries for *The Corrections*, *Purity*, and *Freedom*.

226. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *The Corrections*, one of the Franzen Infringed Works, and titled the infringing and unauthorized derivative "Revisions," using the same characters from Franzen's existing book.

227. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *The Twenty-Seventh City*, one of the Franzen Infringed Works, and titled the infringing and unauthorized derivative “The Rising Metropolis,” using the same characters from Franzen’s existing book.

228. When prompted, ChatGPT generated an accurate summary of the final chapter of *Freedom*, one of the Franzen Infringed Works.

229. ChatGPT could not have generated the material described above if OpenAI’s LLMs had not ingested and been “trained” on the Franzen Infringed Works.

230. **Plaintiff Grisham**. Grisham is a civically engaged and best-selling author. His award-winning work has been translated into approximately 50 languages and adapted for both television and film. Some of Grisham’s most popular novels include *The Pelican Brief*, *The Runaway Jury*, and *The Rainmaker*.

231. Grisham is a member of The Authors Guild.

232. Grisham is the sole author of and owner or beneficial owner of the registered copyrights in twenty-six (26) written works of fiction, all or many of which OpenAI ingested and copied without permission (the “Grisham Infringed Works”).

233. The registration information for the Grisham Infringed Works is contained in Exhibit A to this Complaint, at 5.

234. OpenAI unlawfully and willfully copied the Grisham Infringed Works and used them to “train” OpenAI’s LLMs without Grisham’s permission.

235. For example, when prompted, ChatGPT accurately generated summaries of several of the Grisham Infringed Works, including summaries for *The Chamber*, *The Client*, and *The Firm*.

236. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *The King of Torts*, one of the Grisham Infringed Works, and titled the infringing and unauthorized derivative “The Kingdom of Consequences,” using the same characters from Grisham’s existing book.

237. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *The Last Juror*, one of the Grisham Infringed Works, and titled the infringing and unauthorized derivative “The Juror’s Dilemma,” using the same characters from Grisham’s existing book.

238. When prompted, ChatGPT generated an accurate summary of the final chapter of *The Litigators*, one of the Grisham Infringed Works.

239. ChatGPT could not have generated the material described above if OpenAI’s LLMs had not ingested and been “trained” on the Grisham Infringed Works.

240. **Plaintiff Hilderbrand**. Hilderbrand is a best-selling author, whose works include novels in the romance genre adapted for television. Hilderbrand has previously taught writing at the University of Iowa. Some of Hilderbrand’s most popular novels include *The Summer of ‘69*, *The Identicals*, and *The Perfect Couple*.

241. Hilderbrand is the sole author of and owner or beneficial owner of the registered copyrights in twenty-nine (29) written works of fiction, all or many of which OpenAI ingested and copied without permission (the “Hilderbrand Infringed Works”).

242. The registration information for the Hilderbrand Infringed Works is contained in Exhibit A to this Complaint, at 5–6.

243. OpenAI unlawfully and willfully copied the Hilderbrand Infringed Works and used them “train” OpenAI’s LLMs without Hilderbrand’s permission.

244. For example, when prompted, ChatGPT accurately generated summaries of several of the Hilderbrand Infringed Works, including summaries for *The Summer of '69*, *The Identicals*, and *The Perfect Couple*.

245. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *The Identicals*, one of the Hilderbrand Infringed Works, and titled the infringing and unauthorized derivative “The Reckoning of Twins,” using the same characters from Hilderbrand’s existing book.

246. When prompted, ChatGPT generated an accurate summary of the final chapter of *The Perfect Couple*, one of the Hilderbrand Infringed Works.

247. ChatGPT could not have generated the material described above if OpenAI’s LLMs had not ingested and been “trained” on the Hilderbrand Infringed Works.

248. **Plaintiff Kline**. Kline is a globally published author who writes best-selling novels and has taught different disciplines of writing at Yale University, New York University, and the University of Virginia. Some of Kline’s most popular novels include *Orphan Train*, *A Piece of the World*, and *Bird in Hand*.

249. Kline is a member of The Authors Guild Council and a member of The Authors Guild.

250. Kline is the sole author of and owner or beneficial owner of the registered copyrights in five (5) written works of fiction, all or many of which OpenAI ingested and copied without permission (the “Kline Infringed Works”).

251. The registration information for the Kline Infringed Works is contained in Exhibit A to this Complaint, at 6.

252. OpenAI unlawfully and willfully copied the Kline Infringed Works and used them to “train” OpenAI’s LLMs without Kline’s permission.

253. For example, when prompted, ChatGPT accurately generated summaries of several of the Kline Infringed Works, including summaries for *Orphan Train*, *A Piece of the World*, and *Bird in Hand*.

254. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *Orphan Train*, one of the Kline Infringed Works, and titled the infringing and unauthorized derivative “Legacy Rails,” using the same characters from Kline’s existing book.

255. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *Bird in Hand*, one of the Kline Infringed Works, and titled the infringing and unauthorized derivative “Ties That Bind,” using the same characters from Kline’s existing book.

256. When prompted, ChatGPT generated an accurate summary of the final chapter of *A Piece of the World*, one of the Kline Infringed Works.

257. ChatGPT could not have generated the material described above if OpenAI’s LLMs had not ingested and been “trained” on the Kline Infringed Works.

258. **Plaintiff Lang**. Lang is an author and teacher who holds a doctorate in Comparative Literature. Lang is the author of the novel *The Sixteenth of June*.

259. Lang is the President and a member of The Authors Guild.

260. Lang is the sole author of and owner or beneficial owner of the registered copyrights in one (1) written work of fiction that OpenAI ingested and copied without permission (the “Lang Infringed Work”).

261. The registration information for the Lang Infringed Work is contained in Exhibit A to this Complaint, at 6.

262. OpenAI unlawfully and willfully copied the Lang Infringed Work and used it to “train” OpenAI’s LLMs without Lang’s permission.

263. When prompted, ChatGPT accurately generated a summary of the Lang Infringed Work, *The Sixteenth of June*.

264. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *The Sixteenth of June*, the Lang Infringed Work, and titled the infringing and unauthorized derivative “The Seventeenth of June,” using the same characters from Lang’s existing book.

265. ChatGPT could not have generated the material described above if OpenAI’s LLMs had not ingested and been “trained” on the Lang Infringed Work.

266. **Plaintiff LaValle.** LaValle is an associate professor of Creative Writing at Columbia University and the author of five novels, a short story collection, two novellas, and two comic books. Some of LaValle’s most popular novels include *Big Machine*, *The Devil in Silver*, and *The Changeling*.

267. LaValle is the sole author of and owner or beneficial owner of the registered copyrights in six (6) written works of fiction, all or many of which OpenAI ingested and copied without permission (the “LaValle Infringed Works”).

268. The registration information for the LaValle Infringed Works is contained in Exhibit A to this Complaint, at 6.

269. OpenAI unlawfully and willfully copied the LaValle Infringed Works and used them to “train” OpenAI’s LLMs without LaValle’s permission.

270. For example, when prompted, ChatGPT accurately generated summaries of several of the LaValle Infringed Works, including summaries for *Big Machine*, *The Devil in Silver*, and *The Changeling*.

271. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *The Changeling*, one of the LaValle Infringed Works, and titled the infringing and unauthorized derivative “The Fae’s Return,” using the same characters from LaValle’s existing book.

272. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *The Devil in Silver*, one of the LaValle Infringed Works, and titled the infringing and unauthorized derivative “The New Hyde Legacy,” using the same characters from LaValle’s existing book.

273. When prompted, ChatGPT generated an accurate summary of the final chapter of *Big Machine*, one of the LaValle Infringed Works.

274. ChatGPT could not have generated the material described above if OpenAI’s LLMs had not ingested and been “trained” on the LaValle Infringed Works.

275. **Plaintiff Martin**. Martin is an award-winning author, television producer, and writer who is widely known for his fantasy, science fiction, and horror writing. Some of Martin’s most popular novels include *A Game of Thrones*, *A Clash of Kings*, and *A Storm of Swords*.

276. Martin is the sole author of and owner or beneficial owner of the registered copyrights in fifteen (15) written works of fiction, all or many of which OpenAI ingested and copied without permission (the “Martin Infringed Works”).

277. The registration information for the Martin Infringed Works is contained in Exhibit A to this Complaint, at 6–7.

278. OpenAI unlawfully and willfully copied the Martin Infringed Works and used them to “train” OpenAI’s LLMs without Martin’s permission.

279. In July 2023, Liam Swayne used ChatGPT to generate versions of *The Winds of Winter* and *A Dream of Spring*, intended to be the final two books in the series *A Song of Ice and Fire*, which Martin is currently writing.

280. An experiment conducted by researchers at the University of California, Berkeley, into the “memorization” of works by ChatGPT found that Martin’s novel *A Game of Thrones* ranked 12th with respect to the degree of “memorization.”⁴⁹

281. When prompted, ChatGPT accurately generated summaries of several of the Martin Infringed Works, including summaries for Martin’s novels *A Game of Thrones*, *A Clash of Kings*, and *A Storm of Swords*, the first three books in the series *A Song of Ice and Fire*.

282. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for an alternate sequel to *A Clash of Kings*, one of the Martin Infringed Works, and titled the infringing and unauthorized derivative “A Dance With Shadows,” using the same characters from Martin’s existing books in the series *A Song of Ice and Fire*.

283. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for a prequel book to *A Game of Thrones*, one of the Martin Infringed Works, and titled the infringing and unauthorized derivative “A Dawn of Direwolves,” using the same characters from Martin’s existing books in the series *A Song of Ice and Fire*.

284. When prompted, ChatGPT generated an accurate summary of the final chapter of *The Armageddon Rag*, one of the Martin Infringed Works.

285. ChatGPT could not have generated the results described above if OpenAI’s LLMs had not ingested and been “trained” on the Martin Infringed Works.

⁴⁹ See Kent K. Chang *et al.*, *Speak, Memory: An Archaeology of Books Known to ChatGPT/GPT-4* (2023), available at <https://arxiv.org/pdf/2305.00118v1.pdf> (last accessed Dec. 4, 2023).

286. **Plaintiff Picoult**. A *New York Times* best-selling author, Picoult writes popular fiction. Picoult is also the recipient of many awards, including the New England Bookseller Award for Fiction, the Alex Awards from the YALSA, a lifetime achievement award for mainstream fiction from the Romance Writers of America, the NH Literary Award for Outstanding Literary Merit and the Sarah Josepha Hale Award. Some of Picoult's most popular novels include *My Sister's Keeper*, *Nineteen Minutes*, and *House Rules*.

287. Picoult is a member of The Authors Guild.

288. Picoult is the sole author of and owner or beneficial owner of the registered copyrights in twenty-seven (27) written works of fiction, all or many of which OpenAI ingested and copied without permission (the "Picoult Infringed Works").

289. The registration information for the Picoult Infringed Works is contained in Exhibit A to this Complaint, at 7.

290. OpenAI unlawfully and willfully copied the Picoult Infringed Works and used them to "train" OpenAI's LLMs without Picoult's permission.

291. For example, when prompted, ChatGPT accurately generated summaries of several of the Picoult Infringed Works, including summaries for *Keeping Faith*, *Handle With Care*, and *Sing You Home*.

292. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *The Small Great Things*, one of the Picoult Infringed Works, and titled the infringing and unauthorized derivative "Small Great Things: Unfinished Business," using the same characters from Picoult's existing book.

293. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *My Sister's Keeper*, one of the Picoult Infringed

Works, and titled the infringing and unauthorized derivative as “My Sister’s Legacy,” using the same characters from Picoult’s existing book.

294. When prompted, ChatGPT generated an accurate summary of the final chapter of *Change of Heart*, one of the Picoult Infringed Works.

295. ChatGPT could not have generated the material described above if OpenAI’s LLMs had not ingested and been “trained” on the Picoult Infringed Works.

296. **Plaintiff Preston**. Preston is an author and journalist who has received awards for his writing, both in America and abroad, and previously taught writing at Princeton University. Some of Preston’s most popular novels include *Blasphemy*, *Impact*, and *The Codex*.

297. Preston is a member of The Authors Guild and past President of The Authors Guild Council.

298. Preston is the sole author of and owner or beneficial owner of the registered copyrights in six (6) written works of fiction, all or many of which OpenAI ingested and copied without permission (the “Preston Infringed Works”).

299. The registration information for the Preston Infringed Works is contained in Exhibit A to this Complaint, at 7.

300. OpenAI unlawfully and willfully copied the Preston Infringed Works and used them to “train” OpenAI’s LLMs without Preston’s permission.

301. For example, when prompted, ChatGPT accurately generated summaries of several of the Preston Infringed Works, including summaries for *Impact*, *Blasphemy*, and *The Codex*.

302. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *Impact*, one of the Preston Infringed Works, and

titled the infringing and unauthorized derivative “Unearthed Secrets,” using the same characters from Preston’s existing book.

303. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *The Codex*, one of the Preston Infringed Works, and titled the infringing and unauthorized derivative “The Codex: The Lost Dynasty,” using the same characters from Preston’s existing book.

304. When prompted, ChatGPT generated an accurate summary of the final chapter of *The Kraken Project*, one of the Preston Infringed Works.

305. ChatGPT could not have generated the material described above if OpenAI’s LLMs had not ingested and been “trained” on the Preston Infringed Works.

306. **Plaintiff Robinson.** Robinson is an award-winning author with a wide reach, having written six novels and three collections of short stories, whose fiction has appeared in internationally respected publications and whose books have been published internationally. Some of Robinson’s most popular novels include *Dawson’s Fall*, *Sparta*, and *Cost*.

307. Robinson is a member of The Authors Guild and a past President of The Authors Guild Council.

308. Robinson is the sole author of and owner or beneficial owner of the registered copyrights in eight (8) written works of fiction, all or many of which OpenAI ingested and copied without permission (the “Robinson Infringed Works”).

309. The registration information for the Robinson Infringed Works is contained in Exhibit A to this Complaint, at 7–8.

310. OpenAI unlawfully and willfully copied the Robinson Infringed Works and used them to “train” OpenAI’s LLMs without Robinson’s permission.

311. For example, when prompted, ChatGPT accurately generated summaries of several of the Robinson Infringed Works, including summaries of *Cost*, *Sparta* and *Dawson's Fall*.

312. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *Dawson's Fall*, one of the Robinson Infringed Works, and titled the infringing and unauthorized derivative “Dawson’s Legacy,” using the same characters from Robinson’s existing book.

313. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *Sparta*, one of the Robinson Infringed Works, and titled the infringing and unauthorized derivative “Homefront,” using the same characters from Robinson’s existing book.

314. When prompted, ChatGPT generated an accurate summary of the final chapter of *Sparta*, one of the Robinson Infringed Works.

315. ChatGPT could not have generated the material described above if OpenAI’s LLMs had not ingested and been “trained” on the Robinson Infringed Works.

316. **Plaintiff Saunders**. Saunders is a professor in the English department at Syracuse University, who also writes best-selling books of fiction. Some of Saunders’ most popular works include the short story titled *Escape From Spiderhead*, a novel titled *Lincoln in the Bardo*, and a novella titled *The Brief and Frightening Reign of Phil*.

317. Saunders is a member of The Authors Guild.

318. Saunders is the sole author of and owner or beneficial owner of the registered copyrights in seven (7) written works of fiction, all or many of which OpenAI ingested and copied without permission (the “Saunders Infringed Works”).

319. The registration information for the Saunders Infringed Works is contained in Exhibit A to this Complaint, at 8.

320. OpenAI unlawfully and willfully copied the Saunders Infringed Works and used them to “train” OpenAI’s LLMs without Saunders’s permission.

321. For example, when prompted, ChatGPT accurately generated summaries of several of the Saunders Infringed Works, including summaries for *CivilWarLand in Bad Decline*, *Lincoln in the Bardo*, and *Tenth of December*.

322. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *Fox 8*, one of the Saunders Infringed Works, and titled the infringing and unauthorized derivative “Fox 8 and the Hidden World,” using the same characters from Saunders’s existing book.

323. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *The Tenth of December*, one of the Saunders Infringed Works, and titled the infringing and unauthorized derivative “The Eleventh of December: A Continuation,” using the same characters from Saunders’s existing book.

324. When prompted, ChatGPT generated an accurate summary of the conclusion of *Escape From Spiderhead*, one of the Saunders Infringed Works.

325. ChatGPT could not have generated the material described above if OpenAI’s LLMs had not ingested and been “trained” on the Saunders Infringed Works.

326. **Plaintiff Turow**. A best-selling author, Turow is a novelist and lawyer who is best known for setting his novels in fictional Kindle County’s legal community. Some of Turow’s most popular novels include *The Last Trial*, *Testimony*, and *Identical*.

327. Turow is a member of The Authors Guild and past President of The Authors Guild Council.

328. Turow is the sole author of and owner or beneficial owner of the registered copyrights in sixteen (16) written works of fiction, all or many of which OpenAI ingested and copied without permission (the “Turow Infringed Works”).

329. The registration information for the Turow Infringed Works is contained in Exhibit A to this Complaint, at 8.

330. OpenAI unlawfully and willfully copied the Turow Infringed Works and used them to “train” OpenAI’s LLMs without Turow’s permission.

331. For example, when prompted, ChatGPT accurately generated summaries of several of the Turow Infringed Works, including summaries for *The Burden of Proof*, *Innocent*, and *Testimony*.

332. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *The Last Trial*, one of the Turow Infringed Works, and titled the infringing and unauthorized derivative “Echoes of Judgment,” using the same characters from Turow’s existing book.

333. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *Pleading Guilty*, one of the Turow Infringed Works, and titled the infringing and unauthorized sequel “Redemption’s Price,” using the same characters from Turow’s existing book.

334. When prompted, ChatGPT generated an accurate summary of the final chapter of *Ordinary Heroes*, one of the Turow Infringed Works.

335. ChatGPT could not have generated the material described above if OpenAI’s LLMs had not ingested and been “trained” on the Turow Infringed Works.

336. **Plaintiff Vail**. Rachel Vail is an award-winning American author who primarily authors children's and young adult books. Some of Vail's most popular novels include *Ever After*, *Unfriended*, and *Justin Case: School, Drool, and Other Daily Disasters*.

337. Vail is a member of The Authors Guild and a member of The Authors Guild Council.

338. Vail is the sole author of and owner or beneficial owner of the registered copyrights in twenty-four (24) written works of fiction, all or many of which OpenAI ingested and copied without permission (the "Vail Infringed Works").

339. The registration information for the Vail Infringed Works is contained in Exhibit A to this Complaint, at 8–9.

340. OpenAI unlawfully and willfully copied the Vail Infringed Works and used them to "train" its LLMs without Vail's permission.

341. For example, when prompted, ChatGPT accurately generated summaries of several of the Vail Infringed Works, including summaries for *If We Kiss*, *A Is For Elizabeth*, and *Not That I Care*.

342. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *Bad Best Friend*, one of the Vail Infringed Works, and titled the infringing and unauthorized derivative "Redeeming Friendship," using the same characters from Vail's existing book.

343. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *Do-Over*, one of the Vail Infringed Works, and titled the infringing and unauthorized derivative "Do-Over: Second Chances," using the same characters from Vail's existing book.

344. When prompted, ChatGPT generated an accurate summary of the final chapter of *Daring to be Abigail*, one of the Vail Infringed Works.

345. ChatGPT could not have generated the material described above if OpenAI's LLMs had not ingested and been "trained" on the Vail Infringed Works.

2. Nonfiction Authors

346. While OpenAI and Microsoft have kept the contents of their training data secret, it is likely that, in training their GPT models, they reproduced all or nearly all commercially successful nonfiction books. As OpenAI investor Andreessen Horowitz has admitted, "large language models," like Defendants' GPT models, "are trained on something approaching the entire corpus of the written word," a corpus that would of course include Plaintiffs' works.

347. The size of the Books2 database—the "internet based books corpora" that Defendants used to train GPT-3, GPT-3.5, and possibly GPT-4 as well—has led commentators to believe that Books2 is comprised of books scraped from entire pirated online libraries such as LibGen, ZLibrary, or Bibliotik. Shawn Presser, an independent software developer, created an open-source set of training data called Books3, which was intended to give developers, in his words, "OpenAI-grade training data." The Books3 dataset, similar in size to Books2, was built from a corpus of pirated copies of books available on the site Bibliotik. Works authored and owned by Plaintiffs Alter, Bird, Branch, Cohen, Linden, Okrent, Sancton, Sides, Schiff, Shapiro, Tolentino, and Winchester are available on Books3, an indication that these works were also likely included in the similarly sized Books2.

348. **Plaintiff Alter.** Alter is the author of a number of *New York Times* bestsellers, including *The Center Holds: Obama and His Enemies*; *The Promise: President Obama, Year One*; and *The Defining Moment: FDR's Hundred Days and the Triumph of Hope*. Each of those three books are a part of the Books3 dataset. Pirated copies of each of those three books—as well

as most recent book *His Very Best: Jimmy Carter, A Life*—are available on the internet through websites like LibGen, ZLibrary, and/or Bibliotik, which likely sourced OpenAI's Books2 dataset.

349. When prompted with questions about Alter's books, ChatGPT returned detailed, accurate summaries of them, including of *The Center Holds* and *The Promise*. Upon information and belief, ChatGPT is able to return such detailed information only because it was trained on Plaintiff Alter's books.

350. Alter is the author and owner of the registered copyrights listed under his name in Exhibit B.

351. **Plaintiff Bird**. Bird is the recipient of the 2006 Pulitzer Prize for Biography for *American Prometheus: The Triumph and Tragedy of J. Robert Oppenheimer*. A number of this books, including *American Prometheus*, *The Good Spy*, and *The Color of Truth* are a part of the Books3 dataset. Pirated copies of each of all of his books are available on the internet through websites like LibGen, ZLibrary, and/or Bibliotik, which likely sourced OpenAI's Books2 dataset.

352. When prompted with questions about Bird's books, ChatGPT returned detailed, accurate summaries of them, including of *American Prometheus* and *The Good Spy*. Upon information and belief, ChatGPT is able to return such detailed information only because it was trained on Bird's books.

353. Bird is the author and owner of the registered copyrights listed under his name in Exhibit B.

354. **Plaintiff Branch**. Branch is the author of, among other works, *America in the King Years*, a three-volume history of Martin Luther King Jr. and the Civil Rights Movement. He received the 1989 Pulitzer Prize in History for the first volume in the series, *Parting the Waters*:

America in the King Years, 1954-63. Five of his books, including *Parting the Waters* and *The Clinton Tapes*, are a part of the Books3 dataset. Pirated copies of each of nearly all of his books are available on the internet through websites like LibGen, ZLibrary, and/or Bibliotik, which likely sourced OpenAI's Books2 dataset.

355. When prompted with questions about Branch's books, ChatGPT returned detailed, accurate summaries of them, including of *Parting the Waters* and *The Clinton Tapes*. Upon information and belief, ChatGPT is able to return such detailed information only because it was trained on Plaintiff Branch's books.

356. Branch is the author and owner of the registered copyrights listed under his name in Exhibit B.

357. **Plaintiff Cohen**. Cohen is the author of several *New York Times* bestsellers, including *Tough Jews: Fathers, Sons and Gangster Dreams*; *Sweet and Low: A Family Story*; and *The Sun & the Moon & the Rolling Stones*. He is a contributing editor for *Vanity Fair* and *Rolling Stone* and a columnist for *The Wall Street Journal*.

358. Six of his books, including *Tough Jews* and *The Chicago Cubs*, are a part of the Books3 dataset. Pirated copies of each of nearly all of his books are available on the internet through websites like LibGen, ZLibrary, and/or Bibliotik, which likely sourced OpenAI's Books2 dataset.

359. When prompted with questions about Plaintiff Cohen's books, ChatGPT returned detailed, accurate summaries of them, including of *Tough Jews*. Upon information and belief, ChatGPT is able to return such detailed information only because it was trained on Plaintiff Cohen's books.

360. Cohen is the author and owner of the registered copyrights listed under his name in Exhibit B.

361. **Plaintiff Linden**. Linden is the author of nine nonfiction books, including *The Parrot's Lament, and Other True Tales of Animal Intrigue, Intelligence, and Ingenuity*; *The Octopus and the Orangutan: More True Tales of Animal Intrigue, Intelligence, and Ingenuity*; *The Alms Race: The Impact of American Voluntary Aid Abroad*; and *Winds of Change*.

362. Two of his books, *The Parrot's Lament* and *The Ragged Edge of the World*, are a part of the Books3 dataset. Pirated copies of a number of his books are available on the internet through websites like LibGen, ZLibrary, and/or Bibliotik, which likely sourced OpenAI's Books2 dataset.

363. When prompted with questions about Plaintiff Linden's books, ChatGPT returned detailed, accurate summaries of them, including of *The Octopus and the Orangutan*. Upon information and belief, ChatGPT is able to return such detailed information only because it was trained on Plaintiff Linden's books.

364. Linden is the author and owner of the registered copyrights listed under his name in Exhibit B.

365. **Plaintiff Okrent**. Okrent is the author of a number of nonfiction books, including *Great Fortune: The Epic of Rockefeller Center*, *Last Call: The Rise and Fall of Prohibition*, and *The Guarded Gate: Bigotry, Eugenics and the Law that Kept Two Generations of Jews, Italians and Other European Immigrants Out of America*.

366. At least three of his books, including *Last Call* and *The Guarded Gate*, are a part of the Books3 dataset. Pirated copies of a number of his books are available on the internet through websites like LibGen, ZLibrary, and/or Bibliotik, which likely sourced OpenAI's Books2 dataset.

367. When prompted with questions about Plaintiff Okrent's books, ChatGPT returned detailed, accurate summaries of them, including of *Last Call* and *The Guarded Gate*. Upon

information and belief, ChatGPT is able to return such detailed information only because it was trained on Plaintiff Okrent's books.

368. Okrent is the author and owner of the registered copyrights listed under his name in Exhibit B.

369. **Plaintiff Sancton**. Sancton the author of the *New York Times* bestseller *Madhouse at the End of the Earth: The Belgica's Journey Into the Dark Antarctic Night*. He is a senior features editor of *The Hollywood Reporter* and his work has appeared in *GQ*, *Wired*, and *The New Yorker*.

370. Pirated copies of his book *Madhouse at the End of the Earth* are available on the internet through websites like LibGen, ZLibrary, and/or Bibliotik, which likely sourced OpenAI's Books2 dataset. When prompted with questions about Plaintiff Sancton's book, ChatGPT confirmed that *Madhouse at the End of the Earth* was part of its training dataset.

371. Sancton is the author and owner of the registered copyright listed under his name in Exhibit B.

372. **Plaintiff Sides**. Sides is the author of a number of *New York Times* bestsellers, including *Ghost Soldiers: The Epic Account of World War II's Greatest Rescue Mission* and *Blood and Thunder: An Epic of the American West*. He is an editor-at-large for *Outside* and is a frequent contributor to *National Geographic*.

373. Six of his books, including *Ghost Soldiers* and *Blood and Thunder*, are a part of the Books3 dataset. Pirated copies of all of his books are available on the internet through websites like LibGen, ZLibrary, and/or Bibliotik, which likely sourced OpenAI's Books2 dataset.

374. When prompted with questions about Plaintiff Sides's books, ChatGPT returned detailed, accurate summaries of them, including of *Ghost Soldiers* and *Blood and Thunder*. Upon

information and belief, ChatGPT is able to return such detailed information only because it was trained on Plaintiff Sides's books.

375. Sides is the author and owner of the registered copyrights listed under his name in Exhibit B.

376. **Plaintiff Schiff**. Schiff is the recipient of the 2000 Pulitzer Prize in Biography for *Véra (Mrs. Vladimir Nabokov)*. Plaintiff Schiff is also the author of the *New York Times* bestsellers *Cleopatra: A Life*; *The Witches: Salem, 1692*; and *The Revolutionary: Samuel Adams*. Her work has appeared in *The New Yorker*, *The New York Review of Books*, and *The New York Times*.

377. At least five of her books, including *The Witches* and *Cleopatra*, are a part of the Books3 dataset. Pirated copies of all of her books are available on the internet through websites like LibGen, ZLibrary, and/or Bibliotik, which likely sourced OpenAI's Books2 dataset.

378. When prompted with questions about Plaintiff Schiff's books, ChatGPT returned detailed, accurate summaries of them, including of *The Witches* and *Cleopatra*. Upon information and belief, ChatGPT is able to return such detailed information only because it was trained on Plaintiff Schiff's books.

379. Schiff is the author and owner of the registered copyrights listed under her name in Exhibit B.

380. **Plaintiff Shapiro**. Shapiro is the author of a number of nonfiction books, including *Oberammergau: The Troubling Story of the World's Most Famous Passion Play* and *1599: A Year in the Life of William Shakespeare*. He is the Larry Miller Professor of English and Comparative Literature at Columbia University.

381. At least four of his books, including *The Year of Lear* and *Contested Will*, are a part of the Books3 dataset. Pirated copies of nearly all of his books are available on the internet

through websites like LibGen, ZLibrary, and/or Bibliotik, which likely sourced OpenAI's Books2 dataset.

382. When prompted with questions about Plaintiff Shapiro's books, ChatGPT returned detailed, accurate summaries of them, including of *Contested Will*. Upon information and belief, ChatGPT is able to return such detailed information only because it was trained on Plaintiff Shapiro's books.

383. Shapiro is the author and owner of the registered copyrights listed under his name in Exhibit B.

384. **Plaintiff Tolentino**. She is the author of the *New York Times* bestseller *Trick Mirror: Reflections on Self-Delusion*. She is a staff writer for *The New Yorker* whose work has also appeared in *The New York Times Magazine* and *Pitchfork*.

385. Her book *Trick Mirror* part of the Books3 dataset. Pirated copies of it are widely available on the internet through websites like LibGen, ZLibrary, and/or Bibliotik, which likely sourced OpenAI's Books2 dataset.

386. When prompted with questions about *Trick Mirror*, ChatGPT returned detailed, accurate summaries of the book and its chapters. Upon information and belief, ChatGPT is able to return such detailed information only because it was trained on Plaintiff Tolentino's book.

387. Tolentino is the author and owner of the registered copyright listed under her name in Exhibit B.

388. **Plaintiff Winchester**. Winchester is the author of several *New York Times* bestsellers, including *The Professor and the Madman*; *The Map That Changed the World: William Smith and the Birth of Modern Geology*; and *The Men Who United the States: America's Explorers, Inventors, Eccentrics, and Mavericks, and the Creation of One Nation, Indivisible*.

389. At least eleven of his books, including *The Professor and the Madman* and *Krakatoa*, are a part of the Books3 dataset. Pirated copies of nearly all of his books are available on the internet through websites like LibGen, ZLibrary, and/or Bibliotik, which likely sourced OpenAI's Books2 dataset.

390. When prompted with questions about Plaintiff Winchester's books, ChatGPT returned detailed, accurate summaries of them, including of *The Professor and the Madman* and *Krakatoa*. Upon information and belief, ChatGPT is able to return such detailed information only because it was trained on Plaintiff Winchester's books.

391. Winchester is the author and owner of the registered copyrights listed under his name in Exhibit B.

CLASS ALLEGATIONS

A. Class Definitions

392. This action is brought by Plaintiffs individually and on behalf of the Fiction and Nonfiction Author Classes, as defined below, pursuant to Rule 23(b)(3) and 23(b)(1) of the Federal Rules of Civil Procedure.

393. The Fiction Author Class consists of:

All natural persons in the United States who are the sole authors of, and legal or beneficial owners of Eligible Copyrights in, one or more Fiction Class Works; and all persons in the United States who are the legal or beneficial owners of Eligible Fiction Copyrights in one or more Fiction Class Works held by literary estates.

394. Fiction Class Works is defined as follows:

Any work of fiction, the text of which has been, or is being, used by Defendants to "train" one or more of Defendants' large language models.

395. Eligible Fiction Copyrights is defined as follows:

Any copyright that was registered with the United States Copyright Office before or within five years after first publication of the work, and whose effective date of

registration is either within three months after first publication of the work or before Defendants began using the work to “train” one or more of Defendants’ large language models.

396. The Nonfiction Author Class consists of:

All natural persons, literary trusts, and literary estates in the United States who are legal or beneficial owners of Eligible Nonfiction Copyrights in one or more Nonfiction Class Works; and all persons in the United States who are the legal or beneficial owners of Eligible Nonfiction Copyrights in one or more Fiction Class Works held by literary estates.

397. Nonfiction Class Works is defined as follows:

Any work of nonfiction (A) the text of which has been, or is being, used by Defendants to “train” one or more of Defendants’ large language models; (B) that have been assigned an International Standard Book Number (ISBN); and that fall within a Book Industry Standards and Communications (BISAC) code other than Reference (REF).⁵⁰

398. Eligible Nonfiction Copyrights is defined as follows:

Any copyright that was registered with the United States Copyright Office before or within five years after first publication of the work.

399. Excluded from the definitions of the Fiction Author Class and Nonfiction Author Class above are Defendants; Defendants’ co-conspirators, aiders and abettors, and members of their immediate families; Defendants’ corporate parents, subsidiaries, and affiliates; Defendants’ directors, officers, employees, and other agents, as well as members of their immediate families; and any judge who may preside over this action, the judge’s staff, and members of their immediate families.

B. Rules 23(a) and 23(g)

400. Both Classes consists of at least tens of thousands of authors and copyright holders and thus are so numerous that joinder of all members is impractical.

401. The identities of members of the Classes can be readily ascertained from business records maintained by Defendants.

402. The claims asserted by Plaintiffs are typical of the claims of the Classes because their copyrights were infringed in materially the same way and their interests in preventing future infringement and redressing past infringement are materially the same.

403. The Plaintiffs will fairly and adequately protect the interests of the Classes and do not have any interests antagonistic to those of other members of the Classes.

404. Plaintiffs have retained attorneys who are knowledgeable and experienced in copyright and class action matters, as well as complex litigation.

405. There are questions of fact or law common to the Classes, including:

- a. Whether Defendants' copied works owned by Plaintiffs and the members of the Classes;
- b. Whether Defendants' copying of Plaintiffs' and the Classes' copyrighted works consisted direct, vicarious, or contributory infringement; and
- c. Whether Defendants' copying of works owned by Plaintiffs and the Classes was willful.

C. Rule 23(b)

406. Defendants have acted on grounds common to Plaintiffs and the Classes by treating all Plaintiffs' and Class Members' works equally, in all material respects, in their LLM "training."

407. Common questions of liability for infringement predominate over any individualized damages determinations as may be necessary. To decide liability, the Court will necessarily apply the same law to the same conduct, which Defendants engaged in indiscriminately with respect to all Plaintiffs and all members of the Classes.

⁵⁰ See <http://www.bisg.org/complete-bisac-subject-headings-list> (last visited Feb. 2, 2024)

408. Further, to the extent Plaintiffs elect to pursue statutory rather than actual damages before final judgment, the damages inquiry will likewise be common, if not identical, across Plaintiffs and members of the Classes.

409. A class action is superior to any individual litigation of Plaintiffs' and Class Members' claims. Class Members have little interest, distinct from Plaintiffs' and other Class Members', in prosecuting individual actions. It would waste judicial resources to decide the same legal questions repeatedly, thousands of times over, on materially indistinguishable facts. The Classes presents no special manageability problems.

410. This action is also appropriate as a class action pursuant to Rule 23(b)(2) of the Federal Rules of Civil Procedure because Defendants infringing conduct is applicable generally to Plaintiffs and the proposed Classes and the requested injunctive relief is appropriate respecting the proposed Classes as a whole.

D. Rule 23(c)(4)

411. In the alternative to certification under Rule 23(b)(3) and 23(b)(2), common questions predominate within the determination of liability for infringement, therefore the issue of liability may be separately certified for class treatment even if the entire action is not.

CLAIMS FOR RELIEF

COUNT I: COPYRIGHT INFRINGEMENT (17 U.S.C. § 501)

Against OpenAI and Microsoft

412. Plaintiffs incorporate by reference the allegations in Paragraphs 1 to 411 as though fully set forth herein.

413. Plaintiffs and members of the Classes own the registered copyrights in the works that Defendants reproduced and appropriated to train their artificial intelligence models.

414. Plaintiffs and members of the Classes therefore hold the exclusive rights, including the rights of reproduction and distribution, to those works under 17 U.S.C. § 106.

415. Defendants infringed on the exclusive rights, under 17 U.S.C. § 106, of Plaintiffs and members of the proposed Class by, among other things, reproducing the works owed by Plaintiffs and the proposed Class in datasets used to train their artificial intelligence models.

416. On information and belief, Defendants' infringing conduct alleged herein was and continues to be willful. Defendants infringed on the exclusive rights of Plaintiffs and members of the proposed Class knowing that they were profiting from mass copyright infringement.

417. Plaintiffs and members of the proposed Class are entitled to statutory damages, actual damages, disgorgement, and other remedies available under the Copyright Act.

418. Plaintiffs and members of the proposed Class have been and continue to be irreparably injured due to Defendants' conduct, for which there is no adequate remedy at law. Defendants will continue to infringe on the exclusive right of Plaintiffs and the proposed class unless their infringing activity is enjoined by this Court. Plaintiffs are therefore entitled to permanent injunctive relief barring Defendants' ongoing infringement.

COUNT II: VICARIOUS COPYRIGHT INFRINGEMENT

Against OpenAI Inc. and OpenAI GP LLC

419. Plaintiffs incorporate and reallege paragraphs 1 through 418 above.

420. Defendants OpenAI Inc. and OpenAI GP LLC had the right and ability to control the direct infringement alleged in Count I because Defendant OpenAI Inc. fully controls Defendant OpenAI GP LLC, and Defendant OpenAI GP LLC fully controls Defendant OpenAI OpCo LLC, according to the corporate structure outlined above.

421. Defendants OpenAI Inc. and OpenAI GP LLC have a direct financial interest in the direct infringement alleged in Count I because they benefit from the profits and investments generated by Defendant OpenAI OpCo LLC's infringing activities.

422. Defendants OpenAI Inc. and OpenAI GP LLC are vicariously liable for the direct infringement alleged in Count I.

COUNT III: CONTRIBUTORY INFRINGEMENT

Against Microsoft, OpenAI, Inc., OpenAI GP LLC, OpenAI Global LLC, OpenAI LLC, OAI Corporation LLC, OpenAI Holdings LLC, OpenAI Startup Fund I LP, OpenAI Startup Fund GP I LLC, and OpenAI Startup Fund Management LLC

423. Plaintiffs incorporate by reference and realleges the allegations in Paragraphs 1 to 422 as though fully set forth herein.

424. Microsoft materially contributed and facilitated OpenAI's direct infringement alleged in Count I by providing billions of dollars in investments and designing, creating, and maintaining the bespoke supercomputing system that OpenAI used to maintain and copy the copyrighted works owned by Plaintiffs and the proposed Classes. This assistance was necessary for OpenAI to perpetrate the largescale copyright infringement alleged herein.

425. Microsoft knew, or had reason to know, of the direct infringement alleged in Count I because OpenAI, upon information and belief, informed Microsoft as part of the due diligence process that it was copying and scraping copyrighted material in order to train its generative artificial intelligence models. Furthermore, in the course of designing and maintain its bespoke supercomputing system, Microsoft became aware of OpenAI's direct infringement and directly assisted the copying of copyrighted content owned by Plaintiffs and the proposed Class.

426. Microsoft profited from its OpenAI's direct infringement through its investment in OpenAI and its monetization of GPT-based products.

427. Microsoft is liable for contributing to the direct infringement alleged in Count I.

428. OpenAI, Inc., OpenAI GP LLC, OpenAI Global LLC, OpenAI LLC, OAI Corporation LLC, OpenAI Holdings LLC, OpenAI Startup Fund I LP, OpenAI Startup Fund GP I LLC, and OpenAI Startup Fund Management LLC each directly and indirectly control, direct, and manage other OpenAI entities, including OpenAI OpCo LLC, that are and were responsible for the direct infringement alleged in Count I. These OpenAI entities, through their direction and control of other OpenAI entities, were aware of the direct infringement perpetrated by a variety of OpenAI entities, as alleged in Count I. These OpenAI entities profited from the infringement perpetrated by OpenAI as a whole through their ownership of other OpenAI entities.

429. OpenAI, Inc., OpenAI GP LLC, OpenAI Global LLC, OpenAI LLC, OAI Corporation LLC, and OpenAI Holdings LLC, OpenAI Startup Fund I LP, OpenAI Startup Fund GP I LLC, and OpenAI Startup Fund Management LLC, are each liable for contributing to the direct infringement alleged in Count I.

PRAYER FOR RELIEF

430. Plaintiffs, on behalf of themselves and all others similarly situated, pray for the following relief:

- a. Certification of this action as a class action under Federal Rule of Civil Procedure 23;
- b. Designation of Plaintiffs as class representatives;
- c. Designation of Plaintiffs' counsel as class counsel;
- d. An injunction prohibiting Defendants from infringing on Plaintiffs' and class members' copyrights, including without limitation enjoining Defendants from

using Plaintiffs’ and class members’ copyrighted works in “training” Defendants’ large language models without express authorization;

- e. An award of actual damages to Plaintiffs and class members;
 - f. An award of Defendants’ additional profits attributable to infringement to Plaintiffs and class members;
 - g. An award of statutory damages up to \$150,000 per infringed work to Plaintiffs and members of the Classes, in the alternative to actual damages and profits, at Plaintiffs’ election before final judgment;
- 431. Reasonable attorneys’ fees and costs, as allowed by law;
 - 432. Pre-judgment and post-judgment interest, as allowed by law; and
 - 433. Such further relief as the Court may deem just and proper.

DEMAND FOR JURY TRIAL

Pursuant to Rule 38 of the Federal Rules of Civil Procedure, Plaintiffs hereby demand a jury trial for all claims so triable.

Dated: February 5, 2024

/s/ Rohit D. Nath

Justin A. Nelson (*pro hac vice*)
Alejandra C. Salinas (*pro hac vice*)
SUSMAN GODFREY L.L.P.
1000 Louisiana Street, Suite 5100
Houston, TX 77002
Tel.: 713-651-9366
jnelson@susmangodfrey.com
asalinas@susmangodfrey.com

Rohit D. Nath (*pro hac vice*)
SUSMAN GODFREY L.L.P.
1900 Avenue of the Stars, Suite 1400
Los Angeles, California 90067
Tel.: 310-789-3100
rnath@susmangodfrey.com

J. Craig Smyser
SUSMAN GODFREY L.L.P.
1901 Avenue of the Americas, 32nd Floor
New York, New York 10019
Tel.: 212-336-8330
csmyser@susmangodfrey.com

/s/ Rachel Geman
Rachel Geman
LIEFF CABRASER HEIMANN & BERNSTEIN,
LLP
250 Hudson Street, 8th Floor
New York, New York 10013-1413
Tel.: 212-355-9500
rgeman@lchb.com

Reilly T. Stoler (*pro hac vice forthcoming*)
LIEFF CABRASER HEIMANN & BERNSTEIN,
LLP
275 Battery Street, 29th Floor
San Francisco, CA 94111-3339
Tel.: 415-956-1000
rstoler@lchb.com

Wesley Dozier (*pro hac vice*)
LIEFF CABRASER HEIMANN & BERNSTEIN,
LLP
222 2nd Avenue, Suite 1640
Nashville, TN 37201
Tel.: 615-313-9000
wdozier@lchb.com

/s/ Scott J. Sholder
Scott J. Sholder
CeCe M. Cole
COWAN DEBAETS ABRAHAMS & SHEPPARD
LLP
41 Madison Avenue, 38th Floor
New York, New York 10010
Tel.: 212-974-7474
ssholder@cdas.com
ccole@cdas.com

***Attorneys for Plaintiffs and the Proposed Fiction
and Nonfiction Author Classes***

CERTIFICATE OF SERVICE

I hereby certify this 5th day of February 2024, I caused a true and correct copy of the foregoing to be electronically filed with the Clerk of the court using the CM/ECF system which will send notification to the attorneys of record and is available for viewing and downloading.

/s/ Rohit D. Nath

(Signature)

EXHIBIT A

Registration Number	Title of Work	Author of Work
TX0001472933	Alpine condo crossfire / Mignon G. Eberhart [i.e. Mignon Good Eberhart].	Mignon G. Eberhart
TX0000262270	The Bayou road / Mignon G. Eberhart [i.e. Mignon Good Eberhart].	Mignon G. Eberhart
TX0000464114	Casa Madrone / Mignon G. Eberhart [i.e. Mignon Good Eberhart].	Mignon G. Eberhart
TX0000684927	Family affair / Mignon G. Eberhart [i.e. Mignon Good Eberhart].	Mignon G. Eberhart
TX0001843584	A Fighting chance / Mignon G. Eberhart.	Mignon G. Eberhart
TX0000954810	Next of kin / Mignon G. Eberhart [i.e. Mignon Good Eberhart].	Mignon G. Eberhart
TX0000060127	Nine o'clock tide / Mignon G. Eberhart [i.e. Mignon Good Eberhart].	Mignon G. Eberhart
TX0001195467	The Patient in cabin C / Mignon G. Eberhart [i.e. Mignon Good Eberhart].	Mignon G. Eberhart
TX0002254545	Three days for emeralds / Mignon G. Eberhart.	Mignon G. Eberhart
R167095	The Patient in Room 18	Mignon G. Eberhart
R186351	While the Patient Slept	Mignon G. Eberhart
TX0004447888	Absolute power / David Baldacci.	David Baldacci
TX0008175968	Bullseye: An Original Will Robie / Camel Club Short Story.	David Baldacci
TX0005659939	The Christmas train / David Baldacci.	David Baldacci
TX0006451132	The collectors.	David Baldacci
TX0008977298	Daylight.	David Baldacci
TX0007223564	DELIVER US FROM EVIL.	David Baldacci
TX0007131368	Divine Justice.	David Baldacci
TX0008570267	End Game.	David Baldacci
TX0008577805	The Fallen.	David Baldacci
TX0007877719	The Finisher.	David Baldacci
TX0006936166	First Family.	David Baldacci
TX0008434084	The Fix.	David Baldacci
TX0009074729	A Gambling Man.	David Baldacci
TX0008266478	THE GUILTY.	David Baldacci
TX0007332625	Hell's Corner.	David Baldacci
TX0006109071	Hour game : a novel / David Baldacci	David Baldacci
TX0008151506	The Keeper.	David Baldacci
TX0008429575	The Last Mile.	David Baldacci
TX0008680065	Long Road to Mercy.	David Baldacci
TX0008068212	MEMORY MAN.	David Baldacci
TX0008850402	A Minute to Midnight.	David Baldacci

TX0008388357	No Man's Land.	David Baldacci
TX0007804343	No Time Left.	Columbus Rose, Ltd. f/s/o David Baldacci
TX0008786824	One Good Deed.	David Baldacci
TX0007549427	One Summer.	David Baldacci
TX0008767597	Redemption.	David Baldacci
TX0005099811	Saving faith / David Baldacci.	David Baldacci
TX0006585054	Simple genius.	David Baldacci
TX0004889418	The simple truth / David Baldacci.	David Baldacci
TX0007447606	The Sixth Man.	David Baldacci
TX0005830168	Split second / David Baldacci.	David Baldacci
TX0006939773	STONE COLD.	David Baldacci
TX0007952352	THE TARGET.	David Baldacci
TX0004503942	Total control / David Baldacci.	David Baldacci
TX0007122068	TRUE BLUE.	David Baldacci
TX0008884482	Walk the Wire.	David Baldacci
TX0007106806	The Whole Truth.	David Baldacci
TX0004716435	The winner / David Baldacci.	David Baldacci
TX0006828702	Wish You Well.	David Baldacci
TX0005309086	Wish you well / David Baldacci.	David Baldacci
TX0007446708	Zero Day.	David Baldacci
TX0006811569	DESPERATE DUCHESS.	Mary Bly, whose pseudonym is Eloisa James (author of pseudonymous work)
TX0006522040	Pleasure for pleasure.	Mary Bly (Eloisa James, pseud.)
TX0007238491	A KISS AT MIDNIGHT.	Eloisa James
TX0007916205	WHEN THE DUKE RETURNS.	Eloisa James
TX0008254348	MY AMERICAN DUCHESS.	Eloisa James (author of pseudonymous work)
TX0007044253	A Duke of Her Own.	Eloisa James
TX0005401382	Enchanting pleasures / Eloisa James.	Eloisa James
TX0006969635	This Duchess of Mine.	Eloisa James
TX0005911488	Fool for love / Eloisa James.	Eloisa James
TX0007325306	WHEN BEAUTY TAMED THE BEAST.	Eloisa James
TX0007588681	THE UGLY DUCHESS.	Eloisa James (author of pseudonymous work)
TX0005976986	A wild pursuit / Eloisa James.	Eloisa James
TX0005218280	Potent pleasures / Eloisa James.	Eloisa James
TX0005264332	Midnight pleasures / Eloisa James.	Eloisa James
TX0005727373	Duchess in love / Eloisa James.	Eloisa James

TX0006066736	A fool again / Eloisa James.	Eloisa James
TX0007018439	DUCHESS BY NIGHT.	Eloisa James
TX0005042391	Potent pleasures / Eloisa James.	Eloisa James
TX0006116791	Much ado about you / Eloisa James.	Eloisa James
TX0005979691	Your wicked ways / Eloisa James.	Eloisa James
TX0006297840	Kiss me, Annabel / Eloisa James.	Eloisa James
TX0006354128	The taming of the duke.	Eloisa James
TX0008538194	WILDE IN LOVE : The Wildes of Lindow Castle.	Eloisa James (author of pseudonymous work)
TX0006928439	AN AFFAIR BEFORE CHRISTMAS.	Mary Bly, whose pseudonym is Eloisa James (author of pseudonymous work)
TX0007516011	THE DUKE IS MINE.	Eloisa James, pseud. of Mary Bly (author of pseudonymous work)
TX0007729479	ONCE UPON A TOWER.	Eloisa James, pseud. of Mary Bly (author of pseudonymous work)
TX0008226279	MY AMERICAN DUCHESS.	Eloisa James (author of pseudonymous work)
TX0008888973	SAY YES TO THE DUKE .	Eloisa James (author of pseudonymous work)
TX0008599520	TOO WILDE TO WED.	Eloisa James (author of pseudonymous work)
TX0008499214	SEVEN MINUTES IN HEAVEN.	Eloisa James (author of pseudonymous work)
TX0008768967	SAY NO TO THE DUKE.	Eloisa James (author of pseudonymous work)
TX0007897186	THREE WEEKS WITH LADY X.	Eloisa James (author of pseudonymous work)
TX0008021221	FOUR NIGHTS WITH THE DUKE.	Eloisa James (author of pseudonymous work)
TX0005004510	Angels flight : a novel / Michael Connelly.	Michael Connelly
TX0007697114	The Black Box.	Michael Connelly
TX0003553452	The black ice / by Michael Connelly.	Michael Connelly
TX0004777790	Blood work / Michael Connelly.	Michael Connelly
TX0007487034	The Blue Religion: An Introduction.	Michael Connelly
TX0008966169	The Brass Verdict.	Michael Connelly
TX0006893236	The Brass Verdict.	Michael Connelly
TX0008979084	Chasing the Dime.	Michael Connelly
TX0005698578	Chasing the dime : a novel / by Michael Connelly.	Michael Connelly
TX0006150165	Cielo azul.	Michael Connelly
TX0005511555	City of bones : a novel / by Michael Connelly	Michael Connelly

TX0006218760	The closers.	Michael Connelly
TX0003827545	The concrete blonde / Michael Connelly.	Michael Connelly
TX0006525274	Crime beat : a decade of covering cops and killers.	Michael Connelly
TX0008679767	Dark Sacred Night	Michael Connelly
TX0005244877	A darkness more than night : a novel / by Michael Connelly.	Michael Connelly
TX0007903750	The Drop.	Michael Connelly
TX0007551173	The Drop.	Michael Connelly
TX0006456031	Echo Park.	Michael Connelly
TX0008905639	Fair Warning.	Michael Connelly
TX0007486928	FATHER'S DAY.	Michael Connelly
TX0007377218	The Fifth Witness.	Michael Connelly
TX0007847608	The Gods of Guilt.	Michael Connelly
TX0006921470	In the Shadow of the Master.	Michael Connelly
TX0004064390	The last coyote / Michael Connelly.	Michael Connelly
TX0008943137	The Late Show.	Michael Connelly
TX0008459240	The Late Show.	Michael Connelly
TX0009091556	The Law of Innocence.	Michael Connelly
TX0006979953	THE LINCOLN LAWYER	Michael Connelly
TX0005743159	Lost light : a novel / by Michael Connelly.	Michael Connelly
TX0007000522	MYSTERY WRITERS OF AMERICA PRESENTS BLUE RELIGION: New Stories About Cops, Criminals, and the Chase.	Michael Connelly
TX0006026283	The narrows : a novel / by Michael Connelly	Michael Connelly
TX0008820211	The Night Fire	Michael Connelly
TX0007046637	Nine Dragons	Michael Connelly
TX0006941111	One-Dollar Jackpot.	Michael Connelly
TX0006602700	The overlook.	Michael Connelly
TX0004494924	The poet / Michael Connelly.	Michael Connelly
TX0004290489	The poet / Michael Connelly.	Michael Connelly
TX0007324033	The Reversal.	Michael Connelly
TX0008137420	The Safe Man: A Ghost Story.	Michael Connelly
TX0006998115	The Scarecrow.	Michael Connelly
TX0008154362	Switchblade: An Original Short Story.	Michael Connelly
TX0004571735	Trunk music / Michael Connelly.	Michael Connelly
TX0008530537	Two Kinds of Truth.	Michael Connelly
TX0008387548	The Wrong Side of Goodbye.	Michael Connelly
TX0003234539	The Black echo / Michael Connelly.	Michael Connelly
TX0007571518	BARED TO YOU.	Sylvia Wolicki (AKA Sylvia Day)
TX0008055091	CAPTIVATED BY YOU: A Crossfire Novel.	Sylvia Day

TX0006992994	Don't Tempt Me	Sylvia Day (pseudonym for Sylvia Day Wolicki)
TX0007765389	ENTWINED WITH YOU.	Sylvia Wolicki (AKA Sylvia Day)
TX0007050973	Eve Of Chaos	Sylvia Day writing as S.J. Day (author of pseudonymous work)
TX0007041044	EVE OF DESTRUCTION.	S.J. Day (author of pseudonymous work)
TX0006860949	HEAT OF THE NIGHT.	Sylvia Day
TX0007579762	A HUNGER SO WILD.	Sylvia Wolicki (also known as Sylvia Day)
TX0006853003	Mischief and the Marquess.	Sylvia Day (pseudonym for Sylvia Day Wolicki)
TX0007576661	"On Fire"	Sylvia Wolicki (AKA Sylvia Day)
TX0006848293	A Passion For Him.	Sylvia Day (pseudonym for Sylvia Day Wolicki)
TX0006569840	Pleasures of the night.	Sylvia Day
TX0007310034	Pride and Pleasure.	Sylvia Day, pseud. of Sylvia Day Wolicki
TX0007626414	REFLECTED IN YOU.	Sylvia Wolicki (AKA Sylvia Day)
TX0007426591	Seven Years to Sin.	Sylvia Day, pseud. of Sylvia Day Wolicki
TX0006500604	The stranger I married.	Sylvia Day Wolicki
TX0007706524	"Taking the Heat".	Sylvia Day Wolicki (AKA Sylvia Day)
TX0007487888	A TOUCH OF CRIMSON.	Sylvia Wolicki (also known as Sylvia Day)
TX0007385035	All Revved Up.	Sylvia Day, pseud. of Sylvia June Day
TX0008280509	AFTERBURN.	Sylvia Day, pseud.
TX0007618026	A Dark Kiss of Rapture.	Sylvia Day, pseud. of Sylvia June Wolicki
TX0007620649	What Happened in Vegas.	Sylvia Day, pseud. of Sylvia June Wolicki
TX0006451506	Ask for it.	Sylvia Day (Sylvia Day Wolick, pseud.)
TX0006602177	Passion for the game.	Sylvia Day Wolicki (Sylvia Day, pseud.)

TX0006336892	Bad boys ahoy!	Sylvia Day Wolicki (Sylvia Day, pseud.)
TX0008280508	AFTERSHOCK.	Sylvia Day, pseud.
TX0008788912	Butterfly in Frost.	Sylvia Day, pseud. (author of pseudonymous work)
TX0008817010	CAPTIVATED BY YOU: A Crossfire Novel.	Sylvia Day
TX0008315091	ONE WITH YOU	Sylvia Day
TX0007158134	EVE OF DARKNESS.	S. J. Day (author of pseudonymous work)
TX0006559491	A Familiar Kind of Magic	Sylvia Day
TX0005485729	The corrections / Jonathan Franzen.	Jonathan Franzen
TX0007428625	FREEDOM: A Novel.	Jonathan Franzen
TX0003298778	Strong motion / Jonathan Franzen.	Jonathan Franzen
TX0008137801	PURITY: A Novel.	Jonathan Franzen
TX0002478464	The Twenty-seventh city / Jonathan Franzen.	Jonathan Franzen
TX0005991698	Bleachers	John Grisham
TX0005151784	The brethren / John Grisham.	John Grisham
TX0006096187	The broker / John Grisham.	John Grisham
TX0007543922	CALICO JOE.	John Grisham
TX0003672820	The chamber / John Grisham.	John Grisham
TX0003493490	The client / John Grisham.	John Grisham
TX0003507693	The client / John Grisham.	John Grisham
TX0003021559	The Firm / John Grisham.	John Grisham
TX0007051398	FORD COUNTY: Stories.	John Grisham
TX0008853177	THE GUARDIANS: A Novel.	John Grisham
TXu001050004	The king of torts / by John Grisham.	John Grisham
TX0005706610	The king of torts / John Grisham.	John Grisham
TX0005917373	The last juror / John Grisham.	John Grisham
TX0007494080	THE LITIGATORS.	John Grisham
TX0005349683	A painted house : a novel / by John Grisham	John Grisham
TX0004493634	The partner / John Grisham.	John Grisham
TX0003266464	The Pelican brief / John Grisham.	John Grisham
TX0004069842	The rainmaker / John Grisham.	John Grisham
TX0008691152	THE RECKONING	John Grisham
TX0004331866	The runaway jury / John Grisham	John Grisham
TX0005457447	Skippping Christmas / John Grisham.	John Grisham
TX0004757113	The street lawyer / John Grisham	John Grisham
TX0005477452	The summons / John Grisham.	John Grisham
TX0004945638	The testament / John Grisham.	John Grisham
TX0006991370	A TIME TO KILL.	John Grisham
TX0002655711	A Time to kill / John Grisham.	John Grisham

TX0008882413	28 Summers.	Elin Hilderbrand
TX0006607155	Barefoot.	Elin Hilderbrand
TX0005206228	The beach club / Elin Hilderbrand.	Elin Hilderbrand
TX0007728283	Beautiful Day.	Elin Hilderbrand
TX0006265829	The Blue Bistro.	Elin Hilderbrand
TX0007024786	The Castaways.	Elin Hilderbrand
TX0009013166	Golden Girl.	Elin Hilderbrand
TX0008326741	Here's to Us.	Elin Hilderbrand
TX0008457976	The Identicals.	Elin Hilderbrand
TX0007265117	The Island.	Elin Hilderbrand
TX0006356827	The love season / Elin Hilderbrand.	Elin Hilderbrand
TX0007940979	The Matchmaker.	Elin Hilderbrand
TX0005519951	Nantucket nights / Elin Hilderbrand.	Elin Hilderbrand
TX0008625046	The Perfect Couple.	Elin Hilderbrand
TX0008119893	The Rumor.	Elin Hilderbrand
TX0007436937	Silver Girl.	Elin Hilderbrand
TX0009013724	The Sixth Wedding.	Elin Hilderbrand
TX0006994950	A Summer Affair.	Elin Hilderbrand
TX0008781913	Summer of '69.	Elin Hilderbrand
TX0005717125	Summer people / Elin Hilderbrand.	Elin Hilderbrand
TX0007569862	Summerland.	Elin Hilderbrand
TX0008202402	The Surfing Lesson.	Elin Hilderbrand
TX0008928349	Troubles in Paradise.	Elin Hilderbrand
TX0008821633	What Happens in Paradise.	Elin Hilderbrand
TX0008662697	Winter in Paradise.	Elin Hilderbrand
TX0008671377	Winter Solstice.	Elin Hilderbrand
TX0008324579	Winter Storms.	Elin Hilderbrand
TX0007961716	Winter Street	Elin Hilderbrand
TX0008132271	Winter Stroll	Elin Hilderbrand
TX0007046035	Bird in Hand.	Christina Baker Kline
TX0007705736	ORPHAN TRAIN.	Christina Baker Kline
TX0008384547	A Piece of the World	Christina Baker Kline
TX0006813556	THE WAY LIFE SHOULD BE.	Christina Baker Kline
TX0004971661	Desire lines : a novel	Christina Baker Kline
TX0007909779	The Sixteenth of June.	Maya Lang
TX0007045350	BIG MACHINE: A Novel.	Victor LaValle
TX0008515813	THE CHANGELING A Novel.	Victor LaValle
TX0007637385	THE DEVIL IN SILVER A Novel.	Victor LaValle
TX0009011428	We Travel the Spaceways.	Victor LaValle
TX0005101730	Slapboxing with Jesus : stories / by Victor D. Lavalle.	Victor LaValle

TX0005963757	The ecstatic; or, Homunculus : a novel / Victor LaValle.	Victor LaValle
TX0004947577	A clash of kings : bk. 2 of A song of ice and fire / George R. R. Martin.	George R.R. Martin
TX0007400734	A DANCE WITH DRAGONS Book Five of A Song of Ice and Fire.	George R.R. Martin
TX0006294636	A feast for crows.	George R.R. Martin
TX0001010122	Fevre dream / George R. R. Martin.	George R.R. Martin
TX0008699122	FIRE & BLOOD.	George R.R. Martin
TX0002555019	The Journal of Xavier Desmond.	George R.R. Martin
TX0004387272	A game of thrones / George R. R. Martin.	George R.R. Martin
TX0000672958	The Ice dragon / George R. R. Martin ; ill. by Alicia Austin.	George R.R. Martin
TX0000831294	Sandkings / George R. R. Martin.	George R.R. Martin
TX0008551412	THE SONS OF THE DRAGON.	George R.R. Martin
TX0005306372	A storm of swords / George R. R. Martin.	George R.R. Martin
TX0001847501	Tuf voyaging / George R. R. Martin.	George R.R. Martin
TX0002488534	Winter's chill / by George R. R. Martin.	George R.R. Martin
TX0001359571	The Armageddon rag : a stereophonic long-playing novel / George R. R. Martin.	George R.R. Martin
TX0001314411	Songs the dead men sing / by George R. R. Martin ; ill. by Paul Sonju ; introd. by Algis Budrys.	George R.R. Martin
TX0008908911	THE BOOK OF TWO WAYS A Novel.	Jodi Picoult
TX0007628302	Change of Heart.	Jodi Picoult
TX0007981362	The Color War.	Jodi Picoult
TX0006940734	Handle with Care.	Jodi Picoult
TX0003720035	Harvesting the heart / Jodi Picoult.	Jodi Picoult
TX0007152236	House Rules.	Jodi Picoult
TX0005300595	Keeping faith : a novel / Jodi Picoult.	Jodi Picoult
TX0008760062	LARGER THAN LIFE A Novella.	Jodi Picoult
TX0007948448	LEAVING TIME A Novel.	Jodi Picoult
TX0007509059	Lone Wolf.	Jodi Picoult
TX0005949094	My sister's keeper : a novel / Jodi Picoult.	Jodi Picoult
TX0006542571	Nineteen minutes : a novel / Jodi Picoult.	Jodi Picoult
TX0004708651	The pact : a love story / Jodi Picoult.	Jodi Picoult
TX0005418274	Salem Falls / Jodi Picoult	Jodi Picoult
TX0005714956	Second glance : a novel / Jodi Picoult	Jodi Picoult
TX0008760063	SHINE A Short Story	Jodi Picoult
TX0007343274	SING YOU HOME.	Jodi Picoult
TX0008471033	SMALL GREAT THINGS A Novel	Jodi Picoult
TX0008668131	A SPARK OF LIGHT A Novel	Jodi Picoult
TX0006409872	The tenth circle.	Jodi Picoult

TX0006124529	Vanishing acts.	Jodi Picoult
TX0007210105	Weights & Measures.	Jodi Picoult
TX0004369201	Mercy / Jodi Picoult.	Jodi Picoult
TX0005517659	Perfect match / Jodi Picoult.	Jodi Picoult
TX0004039517	Picture perfect / Jodi Picoult.	Jodi Picoult
TX0005236383	Plain truth / Jodi Picoult.	Jodi Picoult
TX0003356740	Songs of the humpback whale : a novel in five voices / by Jodi Picoult.	Jodi Picoult
TX0006843684	BLASPHEMY.	Douglas Preston
TX0007224608	IMPACT.	Douglas Preston
TX0003921292	Jennie / Douglas Preston.	Douglas Preston
TX0007908862	THE KRAKEN PROJECT.	Douglas Preston
TX0005887183	The codex / Douglas Preston.	Douglas Preston
TX0006206315	Tyrannosaur Canyon.	Douglas Preston
TX0006972227	COST: A Novel.	Roxana Robinson
TX0008776218	Dawson's Fall: A Novel	Roxana Robinson
TX0007796014	SPARTA: A Novel.	Roxana Robinson
TX0002346979	Summer light / Roxana Robinson.	Roxana Robinson
TX0005905727	Sweetwater : a novel / Roxana Robinson.	Roxana Robinson
TX0004867896	This is my daughter : a novel / Roxana Robinson.	Roxana Robinson
TX0007502294	A Perfect Stranger: And Other Stories.	Roxana Robinson
TX0004268621	Asking for love, and other stories / Roxanna Robinson.	Roxana Robinson
TX0006261346	The brief and frightening reign of Phil.	George Saunders
TX0004326821	CivilWarLand in bad decline : stories and a novella / George Saunders	George Saunders
TX0007630065	Escape from Spiderhead.	George Saunders
TX0006374617	In persuasion nation : stories / by George Saunders.	George Saunders
TX0008397114	LINCOLN IN THE BARDO A Novel.	George Saunders
TX0005218791	Pastoralia : stories / by George Saunders.	George Saunders
TX0007666640	TENTH OF DECEMBER Stories.	George Saunders
TX0007815893	Identical.	Scott Turow
TX0007192030	INNOCENT.	Scott Turow
TX0008447808	Testimony.	Scott Turow
TX0002844795	The Burden of proof / Scott Turow.	Scott Turow
TX0004410865	The laws of our fathers / Scott Turow.	Scott Turow
TX0006467762	Limitations.	Scott Turow
TX0006269676	Ordinary heroes.	Scott Turow
TX0005105000	Personal injuries / Scott Turow.	Scott Turow
TX0003554492	Pleading guilty / Scott Turow.	Scott Turow
TX0003554493	Pleading guilty / Scott Turow.	Scott Turow

TX0002110794	Presumed innocent / Scott Turow.	Scott Turow
TX0005655957	Reversible errors / Scott Turow	Scott Turow
TX0006467853	Talking with my mouth full.	Scott Turow
TX0005857010	Ultimate punishment : a lawyer's reflections on dealing with the death penalty / Scott Turow.	Scott Turow
TX0009133293	The Last Trial.	Scott Turow
TX0008929294	The Last Trial.	Scott Turow
TX0009154541	BAD BEST FRIEND.	Rachel Vail
TX0008751369	Big Mouth Elizabeth.	Rachel Vail
TX0007192055	Brilliant.	Rachel Vail
TX0004460895	Daring to be Abigail : a novel / by Rachel Vail.	Rachel Vail
TX0003547708	Do-over : a novel / by Rachel Vail.	Rachel Vail
TX0004009154	Ever after : a novel / by Rachel Vail.	Rachel Vail
TX0005248983	Fill in the blank / Rachel Vail.	Rachel Vail
TX0007041468	Gorgeous.	Rachel Vail
TX0006175909	If we kiss.	Rachel Vail
TX0004839343	If you only knew / Rachel Vail.	Rachel Vail
TX0008750555	A is For Elizabeth.	Rachel Vail
TX0007286754	JUST IN CASE.	Rachel Vail
TX0007889725	JUSTIN CASE: RULES, TOOLS, AND MAYBE A BULLY.	Rachel Vail
TX0007548643	JUSTIN CASE: SHELLS, SMELLS, AND THE HORRIBLE FLIP-FLOPS OF DOOM.	Rachel Vail
TX0007688716	Kiss Me Again.	Rachel Vail
TX0007002999	LUCKY.	Rachel Vail
TX0008001704	UNFRIENDED.	Rachel Vail
TX0008723575	WELL, THAT WAS AWKWARD.	Rachel Vail
TX0005035392	What are friends for? / Rachel Vail.	Rachel Vail
TX0005086489	What are friends for? / Rachel Vail.	Rachel Vail
TX0003243165	Wonder / Rachel Vail.	Rachel Vail
TX0006402673	You, maybe.	Rachel Vail
TX0004916717	Not that I care / Rachel Vail.	Rachel Vail
TX0004916716	Please, please, please / Rachel Vail.	Rachel Vail

Exhibit B

Exhibit B

Plaintiff	Registration Number
Julian Sancton	TX0009331888
Jonathan Alter	TX0007703555
	TX0007189584
	TX0006362219
Kai Bird	TX0006188517
	TX0007857319
	TX0007179572
	TX0004258510
	TX0003294687
Taylor Branch	TX0002701239
	TX0004759338
	TX0006543302
	TX0006997437
	TX0007456285
	TX0007649892
	TX0000910859
Rich Cohen	TX0005050909
	TX0007855718
	TX0006377376
	TX0008311707
	TX0008568164
	TX0009082243
	TX0009184424
	TX0005304760
	TX0005531497
	TX0007607895
	TX0000781640
Eugene Linden	TX0000385052
	TX0001848706
	TX0004843579
	TX0005085858
	TX0005617725
	TX0007369843
	TX0009296965
	TX0000839934
	TX0001573774
Daniel Okrent	TX0002577003
	TX0002689981
	TX0005939955
	TX0006839746
	TX0007180711
	TX0005939955

	TX0005963065
Hampton Sides	TX0005384516
	TX0006449662
	TX0007582071
	TX0007950892
	TX0008698442
	TX0006405264
Stacy Schiff	TX0005225569
	TX0006155631
	TX0007307162
	TX0008713166
	TX0009219559
	TX0005234556
James Shapiro	TX0006274534
	TX0007170652
	TX0008296245
	TX0008907452
	TX0008786905
Jia Tolentino	TX0002031224
Simon Winchester	TX0004848251
	TX0005422703
	TX0006261330
	TX0007786856
	TX0005910779
	TX0007021321
	TX0007303063
	TX0008189441
	TX0008599343

EXHIBIT C

[FILED UNDER SEAL]

EXHIBIT D

[FILED UNDER SEAL]

EXHIBIT E

From: [Malone, Catherine](#)
To: [Stoler, Reilly T.](#); [Seaton, Lauren](#)
Cc: [Kramer, James N.](#); [Dozier, Wesley](#)
Subject: RE: [EXT] Call
Date: Friday, June 14, 2024 1:14:01 PM

Hi Reilly,

Thanks for the call this morning. Below is a summary of what we discussed and the outstanding issues:

Time period: We agree to search for responsive documents from the date our clients began at OpenAI to 6-months after their last day. With respect to direct communications, you clarified that you are asking us to search for direct incoming / outgoing communications with OpenAI and that we would not need to produce any direct communications that are not responsive. We will take that back to our clients.

List of other datasets (Request 5): You will review the protective order for the process of getting us a list of "other datasets," and we will also review the protective order and share it with our clients. We will get back to you on your request that our clients convey the names of other known datasets.

Reasons for departure (Request 1): We offered to provide you a written description of the reasons our clients left OpenAI. You requested that we still produce responsive documents. We will review the protective order language regarding AEO designations and will discuss with our clients.

Depositions: We shared our view that our clients should only be deposed once across the OpenAI cases. We understand that your position is that there is no requirement for you to coordinate depositions with plaintiffs in other actions. However, we also understand that you will endeavor to see what you can offer us in terms of coordination. In the meantime, we will work on figuring out potential dates.

Please let me know if I've misunderstood anything.

We will reach out early next week to schedule our next meet & confer.

Best,
Catherine

From: Stoler, Reilly T. <rstoler@lchb.com>
Sent: Monday, June 10, 2024 4:26 PM
To: Seaton, Lauren <lseaton@orrick.com>
Cc: Malone, Catherine <cmalone@orrick.com>; Kramer, James N. <jkramer@orrick.com>; Dozier, Wesley <wdozier@lchb.com>
Subject: RE: [EXT] Call

[EXTERNAL]

Hi Lauren,

Three follow ups:

1. We can agree to limit all of our requests to the date they began at OpenAI to 6-months after their last day at OpenAI (i.e. sometime in May of 2021), provided you can agree to produce any *direct* correspondence with OAI representatives from their start date to the present.
2. **Re 3(b) below**, we will propose additional dataset search terms (and I will send a preliminary list along) with an understanding that (1) your side will convey the names of relevant, responsive datasets that are known to your clients and (2) that discovery is ongoing and we may need to supplement the list as we obtain discovery from OAI.
3. Can we please set a call for Thursday PM or Friday AM to discuss status/positions on the above and below?

Thanks,

EXHIBIT F

From: [Stoler, Reilly T.](#)
To: ["Seaton, Lauren"; Malone, Catherine](#)
Cc: [Kramer, James N.](#); [Dozier, Wesley](#); [Geman, Rachel](#); [Alter-AI-TT@simplelists.susmangodfrey.com](#)
Subject: RE: [EXT] RE: Amodei/Mann Production
Date: Wednesday, August 28, 2024 6:57:32 PM

Catherine,

Thanks very much for your time today and for addressing our questions about your clients' productions. We understand you have not withheld any documents on the basis of any objection or any instruction from OpenAI.

A few follow ups coming out of our call:

1. It was helpful to understand that the document that hit on the Books2 was not actually referencing the Books2 dataset. As we discussed, please let us know about the Books3 and Children's Book Test dataset hits.
2. You indicated that you had not collected/searched text messages or any other messaging repositories (e.g. WhatsApp, Signal etc.) but would inquire if Messrs. Amodei and Mann have such repositories.

Separately, we discussed scheduling depositions and your side indicated that you would not provide convenient dates for your clients until a deposition coordination order is entered in our proceeding. I indicated that scheduling can be difficult and it would be best to get ahead of the issue but agreed to follow up on the issue if no order is entered in the next week.

Thanks,
Reilly

From: Seaton, Lauren (via Alter-AI-TT list) <Alter-AI-TT@simplelists.susmangodfrey.com>
Sent: Friday, August 23, 2024 7:27 PM
To: Stoler, Reilly T. <rstoler@lchb.com>
Cc: Malone, Catherine <cmalone@orrick.com>; Kramer, James N. <jkramer@orrick.com>; Dozier, Wesley <wdozier@lchb.com>; Geman, Rachel <rgeman@lchb.com>; Alter-AI-TT@simplelists.susmangodfrey.com
Subject: RE: [EXT] RE: Amodei/Mann Production

Hi Reilly,

Would Wednesday, August 28 at 1:30 pm PT work for you?

Thanks,

Lauren

LAUREN SEATON

EXHIBIT G

From: [Stoler, Reilly T.](#)
To: "Malone, Catherine"
Cc: [Kramer, James N.](#); [Dozier, Wesley](#); [Geman, Rachel](#); "Alter-AI-TT@simplelists.susmangodfrey.com"; [Scott, M. Todd](#)
Subject: RE: [EXT] RE: Amodei/Mann - AG v. OAI - Deposition/Production
Date: Tuesday, January 7, 2025 5:10:24 PM
Attachments: [image001.oif](#)

Catherine,

Thanks for the productive call just now.

A few follow ups:

Supplemental Production. You have collected from your clients' relevant messaging repositories—text, whatsapp, and facebook messenger (still awaiting Amodei's FB messenger data). You are preparing a supplemental production of responsive documents identified in these collections. You will provide an update on timing of that production.

Deposition Dates. I noted that discovery closes in April and we'd like to conduct these depositions in February. You agreed to relay convenient deposition dates for Messrs. Amodei and Mann, which will take place in the Bay Area.

Coordination of Mr. Mann's Deposition. You indicated that Mr. Mann was subpoenaed by plaintiffs in the *In re ChatGPT* proceeding and asked if we were open to coordinating our deposition of Mr. Mann with theirs. We're happy to discuss coordination with the *In re ChatGPT* plaintiffs. I previewed that an effective coordinated deposition would likely require that Mr. Mann agree to cross-produce documents in both proceedings. You will follow up to inform us of whether you'd like to move forward with negotiating coordination.

Thanks,

Reilly

From: Malone, Catherine <cmalone@orrick.com>
Sent: Thursday, January 2, 2025 2:56 PM
To: Stoler, Reilly T. <rstoler@lchb.com>
Cc: [Kramer, James N.](#) <jkramer@orrick.com>; [Dozier, Wesley](#) <wdozier@lchb.com>; [Geman, Rachel](#) <rgeman@lchb.com>; 'Alter-AI-TT@simplelists.susmangodfrey.com' <Alter-AI-TT@simplelists.susmangodfrey.com>; [Scott, M. Todd](#) <tscott@orrick.com>
Subject: [EXT] RE: Amodei/Mann - AG v. OAI - Deposition/Production

Hi Reilly,

Since we last spoke, Lauren has switched roles at Orrick and is no longer working on external litigation matters, so I am moving her to BCC.

We are available to meet and confer next Tuesday between 11:30am – 4pm Pacific.

Thanks,
Catherine

From: Stoler, Reilly T. <rstoler@lchb.com>
Sent: Monday, December 30, 2024 11:42 AM

To: Seaton, Lauren <lseaton@orrick.com>; Malone, Catherine <cmalone@orrick.com>
Cc: Kramer, James N. <jkramer@orrick.com>; Dozier, Wesley <wdozier@lchb.com>; Geman, Rachel <rgeman@lchb.com>; 'Alter-AI-TT@simplelists.susmangodfrey.com' <Alter-AI-TT@simplelists.susmangodfrey.com>
Subject: Amodei/Mann - AG v. OAI - Deposition/Production

[EXTERNAL]

Hi Lauren,

Following up on our subpoenas to your clients, please let us know good times to further confer. We would like to discuss (1) scheduling of depositions and (2) confirmation that your clients' have produced all responsive text messages/messaging material.

Thank you,
Reilly



Reilly T. Stoler

Partner

rstoler@lchb.com

t 415.956.1000

f 415.956.1008

Lieff Cabraser Heimann & Bernstein, LLP
275 Battery Street, 29th Floor
San Francisco, CA 94111-3339
www.lieffcabraser.com

This message is intended for the named recipients only. It may contain information protected by the attorney-client or work-product privilege. If you have received this email in error, please notify the sender immediately by replying to this email. Please do not disclose this message to anyone and delete the message and any attachments. Thank you.

NOTICE TO RECIPIENT | This e-mail is meant for only the intended recipient of the transmission, and may be a communication privileged by law. If you received this e-mail in error, any review, use, dissemination, distribution, or copying of this e-mail is strictly prohibited. Please notify us immediately of the error by return e-mail and please delete this message from your system. Thank you in advance for your cooperation.

For more information about Orrick, please visit <http://www.orrick.com>.

In the course of our business relationship, we may collect, store and transfer information about you. Please see our privacy policy at <https://www.orrick.com/Privacy-Policy> to learn about how we use this information.

EXHIBIT H

From: [Malone, Catherine](#)
To: [Stoler, Reilly T.](#)
Cc: [Kramer, James N.](#); [Dozier, Wesley](#); [Geman, Rachel](#); "[Alter-AI-TT@simplelists.susmangodfrey.com](#)"; [Scott, M. Todd](#)
Subject: RE: [EXT] RE: Amodei/Mann - AG v. OAI - Deposition/Production
Date: Thursday, January 16, 2025 5:51:50 PM
Attachments: [image001.gif](#)

Reilly, Wesley, thanks for the call this afternoon.

To follow up on what we discussed:

- Mr. Amodei is not available for deposition.
- Mr. Mann can be available for a four-hour deposition, coordinated with the plaintiffs in the *In re OpenAI ChatGPT Litigation*, and after you complete substantially all of the OpenAI depositions you intend on taking. This offer is contingent on an agreement not to depose Mr. Amodei.

Thanks,
Catherine

From: Stoler, Reilly T. <rstoler@lchb.com>
Sent: Tuesday, January 14, 2025 4:11 PM
To: Malone, Catherine <cmalone@orrick.com>
Cc: Kramer, James N. <jkramer@orrick.com>; Dozier, Wesley <wdozier@lchb.com>; Geman, Rachel <rgeman@lchb.com>; 'Alter-AI-TT@simplelists.susmangodfrey.com' <Alter-AI-TT@simplelists.susmangodfrey.com>; Scott, M. Todd <tscott@orrick.com>
Subject: RE: [EXT] RE: Amodei/Mann - AG v. OAI - Deposition/Production

[EXTERNAL]

Yes, I can confer on Thursday at 1pm.

From: Malone, Catherine <cmalone@orrick.com>
Sent: Tuesday, January 14, 2025 11:24 AM
To: Stoler, Reilly T. <rstoler@lchb.com>
Cc: Kramer, James N. <jkramer@orrick.com>; Dozier, Wesley <wdozier@lchb.com>; Geman, Rachel <rgeman@lchb.com>; 'Alter-AI-TT@simplelists.susmangodfrey.com' <Alter-AI-TT@simplelists.susmangodfrey.com>; Scott, M. Todd <tscott@orrick.com>
Subject: RE: [EXT] RE: Amodei/Mann - AG v. OAI - Deposition/Production

Reilly,

Are you available on Thursday or Friday to further meet and confer? We're available during the following windows:

- Thursday, 10am – 4pm Pacific
- Friday, 10am – 12pm or 2pm – 4pm

Best,
Catherine

From: Stoler, Reilly T. <rstoler@lchb.com>
Sent: Tuesday, January 7, 2025 2:10 PM
To: Malone, Catherine <cmalone@orrick.com>
Cc: Kramer, James N. <jkramer@orrick.com>; Dozier, Wesley <wdozier@lchb.com>; Geman, Rachel <rgeman@lchb.com>; 'Alter-AI-TT@simplelists.susmangodfrey.com' <Alter-AI-TT@simplelists.susmangodfrey.com>; Scott, M. Todd <tscott@orrick.com>
Subject: RE: [EXT] RE: Amodei/Mann - AG v. OAI - Deposition/Production

[EXTERNAL]

Catherine,
Thanks for the productive call just now.
A few follow ups:

Supplemental Production. You have collected from your clients' relevant messaging repositories—text, whatsapp, and facebook messenger (still awaiting Amodei's FB messenger data). You are preparing a supplemental production of responsive documents identified in these collections. You will provide an update on timing of that production.

Deposition Dates. I noted that discovery closes in April and we'd like to conduct these depositions in February. You agreed to relay convenient deposition dates for Messrs. Amodei and Mann, which will take place in the Bay Area.

Coordination of Mr. Mann's Deposition. You indicated that Mr. Mann was subpoenaed by plaintiffs in the *In re ChatGPT* proceeding and asked if we were open to coordinating our deposition of Mr. Mann with theirs. We're happy to discuss coordination with the *In re ChatGPT* plaintiffs. I previewed that an effective coordinated deposition would likely require that Mr. Mann agree to cross-produce documents in both proceedings. You will follow up to inform us of whether you'd like to move forward with negotiating coordination.

Thanks,
Reilly

From: Malone, Catherine <cmalone@orrick.com>
Sent: Thursday, January 2, 2025 2:56 PM
To: Stoler, Reilly T. <rstoler@lchb.com>
Cc: Kramer, James N. <jkramer@orrick.com>; Dozier, Wesley <wdozier@lchb.com>; Geman, Rachel <rgeman@lchb.com>; 'Alter-AI-TT@simplelists.susmangodfrey.com' <Alter-AI-TT@simplelists.susmangodfrey.com>; Scott, M. Todd <tscott@orrick.com>
Subject: [EXT] RE: Amodei/Mann - AG v. OAI - Deposition/Production

Hi Reilly,

Since we last spoke, Lauren has switched roles at Orrick and is no longer working on external litigation matters, so I am moving her to BCC.

We are available to meet and confer next Tuesday between 11:30am – 4pm Pacific.

Thanks,
Catherine

From: Stoler, Reilly T. <rstoler@lchb.com>

Sent: Monday, December 30, 2024 11:42 AM

To: Seaton, Lauren <lseaton@orrick.com>; Malone, Catherine <cmalone@orrick.com>

Cc: Kramer, James N. <jkramer@orrick.com>; Dozier, Wesley <wdozier@lchb.com>; Geman, Rachel <rgeman@lchb.com>; 'Alter-AI-TT@simplelists.susmangodfrey.com' <Alter-AI-TT@simplelists.susmangodfrey.com>

Subject: Amodei/Mann - AG v. OAI - Deposition/Production

[EXTERNAL]

Hi Lauren,

Following up on our subpoenas to your clients, please let us know good times to further confer. We would like to discuss (1) scheduling of depositions and (2) confirmation that your clients' have produced all responsive text messages/messaging material.

Thank you,
Reilly



Reilly T. Stoler

Partner

rstoler@lchb.com

t 415.956.1000

f 415.956.1008

Lieff Cabraser Heimann & Bernstein, LLP

275 Battery Street, 29th Floor

San Francisco, CA 94111-3339

www.lieffcabraser.com

This message is intended for the named recipients only. It may contain information protected by the attorney-client or work-product privilege. If you have received this email in error, please notify the sender immediately by replying to this email. Please do not disclose this message to anyone and delete the message and any attachments. Thank you.

NOTICE TO RECIPIENT | This e-mail is meant for only the intended recipient of the transmission, and may be a communication privileged by law. If you received this e-mail in error, any review, use, dissemination, distribution, or copying of this e-mail is strictly prohibited. Please notify us immediately of the error by return e-mail and please delete this message from your system. Thank you in advance for your cooperation.

For more information about Orrick, please visit <http://www.orrick.com>.

In the course of our business relationship, we may collect, store and transfer information about you. Please see our privacy policy at <https://www.orrick.com/Privacy-Policy> to learn about how we use this information.

This message is intended for the named recipients only. It may contain information protected by the attorney-client or work-product privilege. If you have received this email in error, please notify the sender immediately by replying to this email. Please do not disclose this message to anyone and delete the message and any attachments. Thank you.

NOTICE TO RECIPIENT | This e-mail is meant for only the intended recipient of the transmission, and may be a communication privileged by law. If you received this e-mail in error, any review, use, dissemination, distribution, or copying of this e-mail is strictly prohibited. Please notify us immediately of the error by return e-mail and please delete this message from your system. Thank you in advance for your cooperation.

For more information about Orrick, please visit <http://www.orrick.com>.

In the course of our business relationship, we may collect, store and transfer information about you. Please see our privacy policy at <https://www.orrick.com/Privacy-Policy> to learn about how we use this information.

This message is intended for the named recipients only. It may contain information protected by the attorney-client or work-product privilege. If you have received this email in error, please notify the sender immediately by replying to this email. Please do not disclose this message to anyone and delete the message and any attachments. Thank you.

NOTICE TO RECIPIENT | This e-mail is meant for only the intended recipient of the transmission, and may be a communication privileged by law. If you received this e-mail in error, any review, use, dissemination, distribution, or copying of this e-mail is strictly prohibited. Please notify us immediately of the error by return e-mail and please delete this message from your system. Thank you in advance for your cooperation.

For more information about Orrick, please visit <http://www.orrick.com>.

In the course of our business relationship, we may collect, store and transfer information about you. Please see our privacy policy at <https://www.orrick.com/Privacy-Policy> to learn about how we use this information.

EXHIBIT I

From: [Stoler, Reilly T.](#)
To: ["Malone, Catherine"](#)
Cc: [Kramer, James N.](#); [Dozier, Wesley](#); [Geman, Rachel](#); ["Alter-AI-TT@simplelists.susmangodfrey.com"](#); [Scott, M. Todd](#)
Subject: RE: [EXT] RE: Amodei/Mann - AG v. OAI - Deposition/Production
Date: Friday, January 17, 2025 7:00:57 PM
Attachments: [image001.oif](#)

Catherine + Todd,

Thank you both for your time yesterday. We understand your side's position that Mr. Amodei's work schedule as CEO of Anthropic renders him unavailable for deposition. As we discussed, we disagree. OpenAI has identified Mr. Amodei (and Mr. Mann) as possessing unique, firsthand knowledge of information highly relevant to this case. While we're open to negotiating reasonable accommodations on date, location, duration and coordination with *In re ChatGPT*—as I noted on the call—Mr. Amodei's wholesale refusal to sit for any deposition is a non-starter.

As to Mr. Mann, you've conditioned his availability for a coordinated deposition upon our agreement to forego a deposition of Mr. Amodei (as noted below). For the reasons explained above and detailed on our call, we cannot agree to forego a deposition of Mr. Amodei and therefore understand that Mr. Mann is—like Mr. Amodei—unwilling to sit for deposition.

Please reconsider these positions. As I expressed on our call, I am open to any proposal from your side that would appropriately balance our discovery needs with your clients' constraints. Absent such a proposal, we will proceed formally.

At the conclusion of our call, you asked that we delay any motion to compel Mr. Mann's deposition until after the Court has ruled on his pending motion for a protective order from the deposition subpoena issued in *In re ChatGPT*. To avoid needless motion practice, we can agree to this on the condition that—should the Court deny his current motion—Mr. Mann will not seek another protective order from the deposition subpoena we issued in our proceedings. Please let us know if this agreeable.

Thank you,

Reilly



Reilly T. Stoler

Partner

rstoler@lchb.com

t 415.956.1000

f 415.956.1008

Lieff Cabraser Heimann & Bernstein, LLP
275 Battery Street, 29th Floor
San Francisco, CA 94111-3339
www.lieffcabraser.com

From: Malone, Catherine <cmalone@orrick.com>
Sent: Thursday, January 16, 2025 2:51 PM
To: Stoler, Reilly T. <rstoler@lchb.com>
Cc: Kramer, James N. <jkramer@orrick.com>; Dozier, Wesley <wdozier@lchb.com>; Geman, Rachel <rgeman@lchb.com>; 'Alter-AI-TT@simplelists.susmangodfrey.com' <Alter-AI-TT@simplelists.susmangodfrey.com>; Scott, M. Todd <tscott@orrick.com>
Subject: RE: [EXT] RE: Amodei/Mann - AG v. OAI - Deposition/Production

Reilly, Wesley, thanks for the call this afternoon.

To follow up on what we discussed:

- Mr. Amodei is not available for deposition.
- Mr. Mann can be available for a four-hour deposition, coordinated with the plaintiffs in the *In re OpenAI ChatGPT Litigation*, and after you complete substantially all of the OpenAI depositions you intend on taking. This offer is contingent on an agreement not to depose Mr. Amodei.

Thanks,
Catherine

From: Stoler, Reilly T. <rstoler@lchb.com>
Sent: Tuesday, January 14, 2025 4:11 PM
To: Malone, Catherine <cmalone@orrick.com>
Cc: Kramer, James N. <jkramer@orrick.com>; Dozier, Wesley <wdozier@lchb.com>; Geman, Rachel <rgeman@lchb.com>; 'Alter-AI-TT@simplelists.susmangodfrey.com' <Alter-AI-TT@simplelists.susmangodfrey.com>; Scott, M. Todd <tscott@orrick.com>
Subject: RE: [EXT] RE: Amodei/Mann - AG v. OAI - Deposition/Production

[EXTERNAL]

Yes, I can confer on Thursday at 1pm.

From: Malone, Catherine <cmalone@orrick.com>
Sent: Tuesday, January 14, 2025 11:24 AM
To: Stoler, Reilly T. <rstoler@lchb.com>
Cc: Kramer, James N. <jkramer@orrick.com>; Dozier, Wesley <wdozier@lchb.com>; Geman, Rachel <rgeman@lchb.com>; 'Alter-AI-TT@simplelists.susmangodfrey.com' <Alter-AI-TT@simplelists.susmangodfrey.com>; Scott, M. Todd <tscott@orrick.com>
Subject: RE: [EXT] RE: Amodei/Mann - AG v. OAI - Deposition/Production

Reilly,

Are you available on Thursday or Friday to further meet and confer? We're available during the following windows:

- Thursday, 10am – 4pm Pacific

Friday, 10am – 12pm or 2pm – 4pm

Best,
Catherine

From: Stoler, Reilly T. <rstoler@lchb.com>
Sent: Tuesday, January 7, 2025 2:10 PM
To: Malone, Catherine <cmalone@orrick.com>
Cc: Kramer, James N. <jkramer@orrick.com>; Dozier, Wesley <wdozier@lchb.com>; Geman, Rachel <rgeman@lchb.com>; 'Alter-AI-TT@simplelists.susmangodfrey.com' <Alter-AI-TT@simplelists.susmangodfrey.com>; Scott, M. Todd <tscott@orrick.com>
Subject: RE: [EXT] RE: Amodei/Mann - AG v. OAI - Deposition/Production

EXTERNAL

Catherine,
Thanks for the productive call just now.
A few follow ups:

Supplemental Production. You have collected from your clients' relevant messaging repositories—text, whatsapp, and facebook messenger (still awaiting Amodei's FB messenger data). You are preparing a supplemental production of responsive documents identified in these collections. You will provide an update on timing of that production.

Deposition Dates. I noted that discovery closes in April and we'd like to conduct these depositions in February. You agreed to relay convenient deposition dates for Messrs. Amodei and Mann, which will take place in the Bay Area.

Coordination of Mr. Mann's Deposition. You indicated that Mr. Mann was subpoenaed by plaintiffs in the *In re ChatGPT* proceeding and asked if we were open to coordinating our deposition of Mr. Mann with theirs. We're happy to discuss coordination with the *In re ChatGPT* plaintiffs. I previewed that an effective coordinated deposition would likely require that Mr. Mann agree to cross-produce documents in both proceedings. You will follow up to inform us of whether you'd like to move forward with negotiating coordination.

Thanks,
Reilly

From: Malone, Catherine <cmalone@orrick.com>
Sent: Thursday, January 2, 2025 2:56 PM
To: Stoler, Reilly T. <rstoler@lchb.com>
Cc: Kramer, James N. <jkramer@orrick.com>; Dozier, Wesley <wdozier@lchb.com>; Geman, Rachel <rgeman@lchb.com>; 'Alter-AI-TT@simplelists.susmangodfrey.com' <Alter-AI-TT@simplelists.susmangodfrey.com>; Scott, M. Todd <tscott@orrick.com>
Subject: [EXT] RE: Amodei/Mann - AG v. OAI - Deposition/Production

Hi Reilly,

Since we last spoke, Lauren has switched roles at Orrick and is no longer working on external litigation matters, so I am moving her to BCC.

We are available to meet and confer next Tuesday between 11:30am – 4pm Pacific.

Thanks,
Catherine

From: Stoler, Reilly T. <rstoler@lchb.com>

Sent: Monday, December 30, 2024 11:42 AM

To: Seaton, Lauren <lseaton@orrick.com>; Malone, Catherine <cmalone@orrick.com>

Cc: Kramer, James N. <jkramer@orrick.com>; Dozier, Wesley <wdozier@lchb.com>; Geman, Rachel <rgeman@lchb.com>; 'Alter-AI-TT@simplelists.susmangodfrey.com' <Alter-AI-TT@simplelists.susmangodfrey.com>

Subject: Amodei/Mann - AG v. OAI - Deposition/Production

[EXTERNAL]

Hi Lauren,

Following up on our subpoenas to your clients, please let us know good times to further confer. We would like to discuss (1) scheduling of depositions and (2) confirmation that your clients' have produced all responsive text messages/messaging material.

Thank you,
Reilly



Reilly T. Stoler

Partner

rstoler@lchb.com

t 415.956.1000

f 415.956.1008

Lieff Cabraser Heimann & Bernstein, LLP
275 Battery Street, 29th Floor
San Francisco, CA 94111-3339
www.lieffcabraser.com

This message is intended for the named recipients only. It may contain information protected by the attorney-client or work-product privilege. If you have received this email in error, please notify the sender immediately by replying to this email. Please do not disclose this message to anyone and delete the message and any attachments. Thank you.

NOTICE TO RECIPIENT | This e-mail is meant for only the intended recipient of the transmission, and may be a communication privileged by law. If you received this e-mail in error, any review, use, dissemination, distribution, or copying of this e-mail is strictly prohibited. Please notify us immediately of the error by return e-mail and please delete this message from your system. Thank you in advance for your cooperation.

For more information about Orrick, please visit <http://www.orrick.com>.

In the course of our business relationship, we may collect, store and transfer information about you. Please see our privacy policy at <https://www.orrick.com/Privacy-Policy> to learn about how we use this information.

This message is intended for the named recipients only. It may contain information protected by the attorney-client or work-product privilege. If you have received this email in error, please notify the sender immediately by replying to this email. Please do not disclose this message to anyone and delete the message and any attachments. Thank you.

NOTICE TO RECIPIENT | This e-mail is meant for only the intended recipient of the transmission, and may be a communication privileged by law. If you received this e-mail in error, any review, use, dissemination, distribution, or copying of this e-mail is strictly prohibited. Please notify us immediately of the error by return e-mail and please delete this message from your system. Thank you in advance for your cooperation.

For more information about Orrick, please visit <http://www.orrick.com>.

In the course of our business relationship, we may collect, store and transfer information about you. Please see our privacy policy at <https://www.orrick.com/Privacy-Policy> to learn about how we use this information.

This message is intended for the named recipients only. It may contain information protected by the attorney-client or work-product privilege. If you have received this email in error, please notify the sender immediately by replying to this email. Please do not disclose this message to anyone and delete the message and any attachments. Thank you.

NOTICE TO RECIPIENT | This e-mail is meant for only the intended recipient of the transmission, and may be a communication privileged by law. If you received this e-mail in error, any review, use, dissemination, distribution, or copying of this e-mail is strictly prohibited. Please notify us immediately of the error by return e-mail and please delete this message from your system. Thank you in advance for your cooperation.

For more information about Orrick, please visit <http://www.orrick.com>.

In the course of our business relationship, we may collect, store and transfer information about you. Please see our privacy policy at <https://www.orrick.com/Privacy-Policy> to learn about how we use this information.

EXHIBIT J

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

AUTHORS GUILD, et al.,

Plaintiffs,

v.

OPEN AI INC., et al.,

Defendants.

ECF CASE

No. 1:23-cv-08292-SHS;

No. 1:23-cv-10211-SHS

**PLAINTIFFS' FIRST SET OF
INTERROGATORIES TO
OPENAI**

JONATHAN ALTER, et al.,

Plaintiffs,

v.

OPENAI, INC., et al.,

Defendants.

PLAINTIFFS' FIRST SET OF INTERROGATORIES TO OPENAI

Plaintiffs, by and through their undersigned attorneys, request that OpenAI provide an answer to the Interrogatories within thirty (30) days of the date of service hereof as provided by Federal Rule of Civil Procedure 33.

DEFINITIONS

1. “**Action**” shall mean *Authors Guild, et al., v. OpenAI, et al.*, No. 1:23-cv-08292-SHS (S.D.N.Y.) and *Jonathan Alter, et al., v. Open AI Inc. et al.*, No. 1:23-cv-10211-SHS (S.D.N.Y.).

2. “**Agreement(s)**” means any writing or record of any type or description known to You including, but not limited to, any existing contract, contract renewal, letter of commitment,

addendum, draft, revision, amendment, renewal, and settlements, that effectuates or otherwise impacts a business relationship between You and another Person. If any interrogatory asks “whether there are any Agreements,” please identify all such Agreements by Bates number. If any such Agreement has not been produced, please produce it and identify it by Bates number. If there is no such Agreement, please state so.

3. **“Communication(s)”** means the transmittal of information (in the form of facts, ideas, inquiries or otherwise) by any means, including, but not limited to, telephone calls, emails (whether via company server or personal webmail or similar accounts), faxes, text messages (on work or personal phones), instant messages, Skype, Line, WhatsApp, WeChat, other electronic messages, letters, notes, and voicemails.

4. **“Date”** means the exact date, month, and year, if ascertainable or, if not, the best approximation of the date (based upon relationship with other events). If the date provided is an approximation, please so state.

5. **“Document(s)”** is defined to be synonymous in meaning and equal in scope to the usage of the term “documents or electronically stored information” in Federal Rule of Civil Procedure 34(a)(1)(A).

6. **“ESI”** refers to information and Documents within the full scope of Federal Rule of Civil Procedure 34 with all metadata intact—created, manipulated, communicated, stored, and best utilized in digital form, and stored on electronic media. Examples of ESI include e-mail, messages posted on electronic message boards, forum postings, support tickets, videos, discussion boards, data, source code, websites, Microsoft Word files, Microsoft Excel files, and instant messages.

7. **“Investors”** means any person or entity that OpenAI has approached through

written and/or oral communication to discuss a potential financial investment in the operations and/or entitie(s) controlled and/or overseen by one or more of the OpenAI defendant entities.

8. **“LLMs”** has the same meaning as the term discussed in OpenAI’s letter to the Register of Copyrights and Director of the U.S. Copyright Office dated October 30, 2023, “Re: Notice of Inquiry and Request for Comment [Docket No. 2023-06]”.

9. **“Microsoft”** means Microsoft Corporation, as well as corporate parents, subsidiaries, affiliates, predecessor entities, successor entities, divisions, departments, groups, acquired entities, related entities, or any other entity acting or purporting to act on their behalf.

10. **“Microsoft Individual(s)”** means all present and former directors, officers, employees, agents, attorneys, consultants and representatives of Microsoft Corporation; and all Persons either acting or purporting to act on behalf of Microsoft Corporation.

11. **“OpenAI Individual(s)”** means all present and former directors, officers, employees, agents, attorneys, consultants and representatives of OpenAI, Inc., OpenAI GP, LLC, OpenAI, LLC, OpenAI OPCO LLC, OpenAI Global LLC, OAI Corporation, LLC, and OpenAI Holdings, LLC; and all Persons either acting or purporting to act on behalf of OpenAI, Inc., OpenAI GP, LLC, OpenAI, LLC, OpenAI OPCO LLC, OpenAI Global LLC, and OAI Corporation, LLC, OpenAI Holdings, LLC.

12. **“Person(s)”** means any natural person.

13. **“You”, “Your”, and “OpenAI”** means OpenAI, Inc., OpenAI GP, LLC, OpenAI, LLC, OpenAI OPCO LLC, OpenAI Global LLC, OAI Corporation, LLC, OpenAI Holdings, LLC, as well as corporate parents, subsidiaries, affiliates, predecessor entities, successor entities, divisions, departments, groups, acquired entities, related entities, or any other entity acting or purporting to act on their behalf.

INTERROGATORIES

INTERROGATORY NO. 1:

Identify the five OpenAI Individuals who are most knowledgeable about the creation and contents of the Books2 dataset. For each OpenAI Individual identified, include (i) the Person's full name; (ii) any position the Person held with You at the time the Person possessed this knowledge; (iii) any email addresses the Person used to communicate about Books2; and (iv) the name of any other form of ESI (e.g., text messages, Teams, and Slack) the Person used to communicate about or otherwise document information related to Books2.

INTERROGATORY NO. 2:

Identify the five OpenAI Individuals who are most knowledgeable about Your data retention policies. For each OpenAI Individual identified, include (i) the Person's full name; (ii) any position the Person held with You at the time the Person possessed this knowledge; (iii) any email addresses the Person used to communicate about the retention policies; and (iv) the name of any other form of ESI (e.g., text messages, Teams, and Slack) the Person used to communicate about or otherwise document information related to the retention policies.

INTERROGATORY NO. 3:

Identify any Investors You approached about a potential investment in any OpenAI entity and/or Your LLMs.

INTERROGATORY NO. 4:

Identify Microsoft Individuals with whom You have discussed licensing Agreements for the use of copyrighted materials.

INTERROGATORY NO. 5:

Identify any third-party individuals or organizations with which You have discussed licensing Agreements for any copyrighted materials.

INTERROGATORY NO. 6:

Identify any third-party individuals or organizations with which you have entered into any licensing Agreements for the use of copyrighted materials.

INTERROGATORY NO. 7:

Identify the five OpenAI Individuals who are most knowledgeable about Your training dataset used to train LLMs. For each OpenAI Individual identified, include (i) the Person's full name; (ii) any position the Person held with You at the time the Person possessed this knowledge; (iii) any email addresses the Person used to communicate about the training dataset; and (iv) the name of any other form of ESI (e.g., text messages, Teams, and Slack) the Person used to communicate about or otherwise document information related to the training dataset.

Dated: February 23, 2024

/s/ Justin A. Nelson

Justin A. Nelson (*pro hac vice*)
Alejandra C. Salinas (*pro hac vice*)
SUSMAN GODFREY L.L.P.
1000 Louisiana Street, Suite 5100
Houston, TX 77002
Tel.: 713-651-9366
jnelson@susmangodfrey.com
asalinas@susmangodfrey.com

Rohit D. Nath (*pro hac vice*)
SUSMAN GODFREY L.L.P.
1900 Avenue of the Stars, Suite 1400
Los Angeles, California 90067
Tel.: 310-789-3100
rnath@susmangodfrey.com

J. Craig Smyser
SUSMAN GODFREY L.L.P.
1901 Avenue of the Americas, 32nd Floor
New York, New York 10019
Tel.: 212-336-8330
csmyser@susmangodfrey.com

/s/ Rachel German
Rachel Geman
LIEFF CABRASER HEIMANN & BERNSTEIN,
LLP
250 Hudson Street, 8th Floor
New York, New York 10013-1413
Tel.: 212-355-9500
rgeman@lchb.com

Reilly T. Stoler (*pro hac vice forthcoming*)
Ian M. Bensberg (*pro hac vice*)
LIEFF CABRASER HEIMANN & BERNSTEIN,
LLP
275 Battery Street, 29th Floor
San Francisco, CA 94111-3339
Tel.: 415-956-1000
rstoler@lchb.com
ibenserg@lchb.com

Wesley Dozier (*pro hac vice*)
LIEFF CABRASER HEIMANN & BERNSTEIN,
LLP
222 2nd Avenue, Suite 1640
Nashville, TN 37201
Tel.: 615-313-9000
wdozier@lchb.com

/s/ Scott J. Sholder
Scott J. Shoulder
CeCe M. Cole
COWAN DEBAETS ABRAHAMS & SHEPPARD
LLP
41 Madison Avenue, 38th Floor
New York, New York 10010
Tel.: 212-974-7474
ssholder@cdas.com
ccole@cdas.com

Attorney for Plaintiffs and the Proposed Class

CERTIFICATE OF SERVICE

I hereby certify that on February 23, 2024, a copy of the foregoing was served via electronic mail to all counsel of record in this matter.

/s/ Alejandra C. Salinas
Alejandra C. Salinas

EXHIBIT K

[FILED UNDER SEAL]